

VOLUME 355 NUMBER 7



JULY 20

WHOLE NUMBER 3

TRANSACTIONS

OF THE

112

A M E R I C A N M A T H E M A T I C A L S O C I E T Y

EDITED BY

Dan Abramovich

Peter W. Bates

Patricia E. Bauman

William Beckner, Managing Editor

Krzysztof Burdzy

Tobias Colding

Harold G. Diamond

Sergey Fomin

Lisa C. Jeffrey

Alexander Nagel

D. H. Phong

Stewart Priddy

Theodore Slaman

Karen E. Smith

Robert J. Stanton

Daniel Tataru

Abigail Thompson

Robert F. Williams

PROVIDENCE, RHODE ISLAND USA

ISSN 0002-9947

Available electronically

Transactions of the American Mathematical Society

This journal is devoted entirely to research in pure and applied mathematics.

Submission information. See **Information for Authors** at the end of this issue.

Publisher Item Identifier. The Publisher Item Identifier (PII) appears at the top of the first page of each article published in this journal. This alphanumeric string of characters uniquely identifies each article and can be used for future cataloging, searching, and electronic retrieval.

Postings to the AMS website. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

Subscription information. *Transactions of the American Mathematical Society* is published monthly. Beginning in January 1996 *Transactions* is accessible from www.ams.org/publications/. Subscription prices for Volume 355 (2003) are as follows: for paper delivery, \$1490 list, \$1192 institutional member, \$1341 corporate member; for electronic delivery, \$1341 list, \$1073 institutional member, \$1207 corporate member. Upon request, subscribers to paper delivery of this journal are also entitled to receive electronic delivery. If ordering the paper version, add \$39 for surface delivery outside the United States and India; \$50 to India. Expedited delivery to destinations in North America is \$48; elsewhere \$144. For paper delivery a late charge of 10% of the subscription price will be imposed upon orders received from nonmembers after January 1 of the subscription year.

Back number information. For back issues see www.ams.org/bookstore.

Subscriptions and orders should be addressed to the American Mathematical Society, P.O. Box 845904, Boston, MA 02284-5904 USA. *All orders must be accompanied by payment.* Other correspondence should be addressed to 201 Charles Street, Providence, RI 02904-2294 USA.

Copying and reprinting. Material in this journal may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. Requests can also be made by e-mail to reprint-permission@ams.org.

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

Transactions of the American Mathematical Society is published monthly by the American Mathematical Society at 201 Charles Street, Providence, RI 02904-2294 USA. Periodicals postage is paid at Providence, Rhode Island. Postmaster: Send address changes to *Transactions*, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

© 2003 by the American Mathematical Society. All rights reserved.

This journal is indexed in *Mathematical Reviews*, *Zentralblatt MATH*, *Science Citation Index*®, *Science Citation Index*™-Expanded, *ISI Alerting Services*™, *CompuMath Citation Index*®, and *Current Contents*®/Physical, Chemical & Earth Sciences.

Printed in the United States of America.

⊗ The paper used in this journal is acid-free and falls within the guidelines established to ensure permanence and durability.

TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY

CONTENTS

Vol. 355, No. 7

Whole No. 818

July 2003

Borislav Karaivanov, Pencho Petrushev, and Robert C. Sharpley, Algorithms for nonlinear piecewise polynomial approximation: Theoretical aspects	2585
Jörg Brendle, The almost-disjointness number may have countable cofinality	2633
Alina Carmen Cojocaru, Cyclicity of CM elliptic curves modulo p	2651
Tonghai Yang, Taylor expansion of an Eisenstein series	2663
Eric Freeman, Systems of diagonal Diophantine inequalities	2675
Francisco Javier Gallego and Bangere P. Purnaprajna, On the canonical rings of covers of surfaces of minimal degree	2715
H. H. Brungs and N. I. Dubrovin, A classification and examples of rank one chain domains	2733
Donald W. Barnes, On the spectral sequence constructors of Guichardet and Stefan	2755
Steven Lillywhite, Formality in an equivariant setting	2771
Neil Hindman, Dona Strauss, and Yevhen Zelenyuk, Large rectangular semigroups in Stone-Čech compactifications	2795
Takehiko Yamanouchi, Galois groups of quantum group actions and regularity of fixed-point algebras	2813
Boo Rim Choe, Hyungwoon Koo, and Wayne Smith, Composition operators acting on holomorphic Sobolev spaces	2829
B. Jakubczyk and M. Zhitomirskii, Distributions of corank 1 and their characteristic vector fields	2857
E. Boeckx, When are the tangent sphere bundles of a Riemannian manifold reducible?	2885
Henri Comman, Criteria for large deviations	2905
Seung Jun Chang, Jae Gil Choi, and David Skoug, Integration by parts formulas involving generalized Fourier-Feynman transforms on function space	2925
Michiko Yuri, Thermodynamic formalism for countable to one Markov systems	2949
D. G. De Figueiredo and Y. H. Ding, Strongly indefinite functionals and multiple solutions of elliptic systems	2973
F. Rousset, Stability of small amplitude boundary layers for mixed hyperbolic-parabolic systems	2991



ALGORITHMS FOR NONLINEAR PIECEWISE POLYNOMIAL APPROXIMATION: THEORETICAL ASPECTS

BORISLAV KARAIVANOV, PENCHO PETRUSHEV, AND ROBERT C. SHARPLEY

ABSTRACT. In this article algorithms are developed for nonlinear n -term Courant element approximation of functions in L_p ($0 < p \leq \infty$) on bounded polygonal domains in \mathbb{R}^2 . Redundant collections of Courant elements, which are generated by multilevel nested triangulations allowing arbitrarily sharp angles, are investigated. Scalable algorithms are derived for nonlinear approximation which both capture the rate of the best approximation and provide the basis for numerical implementation. Simple thresholding criteria enable approximation of a target function f to optimally high asymptotic rates which are determined and automatically achieved by the inherent smoothness of f . The algorithms provide direct approximation estimates and permit utilization of the general Jackson-Bernstein machinery to characterize n -term Courant element approximation in terms of a scale of smoothness spaces (B -spaces) which govern the approximation rates.

1. INTRODUCTION

Highly detailed Digital Terrain Elevation Data (DTED) and associated imagery are now becoming widely available for most of the earth's surface. However, algorithms for effective approximation of data of this type are not yet available. A primary motivation for this work is the development of effective algorithms for nonlinear piecewise polynomial approximation of DTED maps from a redundant hierarchial system over (possibly) irregular triangulations which are constructive in nature. Application of the ideas and theory from [4] to the resulting framework will permit optimal entropy tree encoding of the elevation data, enable progressive view-dependent refinements which may be focused to user-localized regions, and permit the registration of similarly encoded image textures to the surface (see [10], [4] for more details).

Our philosophy is that dependable practical approximation procedures can be built only upon a solid theoretical basis. Accordingly, we have two primary goals in this paper. The first is to better understand nonlinear piecewise polynomial

Received by the editors May 2, 2002.

2000 *Mathematics Subject Classification.* Primary 41A17, 41A25, 65D18; Secondary 65D07, 42B35.

Key words and phrases. Nested irregular triangulations, redundant representations, nonlinear n -term approximation, Courant elements, Jackson and Bernstein estimates.

The second and third authors were supported in part by Grant NSF #DMS-0079549 and ONR N00014-01-1-0515.

All three authors were supported in part by ONR grant N00014-00-1-0470.

approximation, in particular, to understand the nature of the global smoothness conditions (spaces) which govern the rate of approximation. The second goal is to develop or refine existing constructive approximation methods for nonlinear approximation which capture the rate of the best approximation and can be implemented effectively in practice.

This paper addresses nonlinear n -term approximation by Courant elements generated by multilevel nested triangulations. More precisely, for a given bounded polygonal domain $E \subset \mathbb{R}^2$, let $(\mathcal{T}_m)_{m \geq 0}$ be a sequence of triangulations such that each level \mathcal{T}_m is a triangulation of E consisting of closed triangles with disjoint interiors and a refinement of the previous level \mathcal{T}_{m-1} . We impose some mild natural conditions on the triangulations in order to prevent possible deterioration, but our results are valid for fairly general triangulations with sharp angles. We define $\mathcal{T} := \bigcup_{m \geq 0} \mathcal{T}_m$. Each such multilevel triangulation \mathcal{T} generates a ladder of spaces $\mathcal{S}_0 \subset \mathcal{S}_1 \subset \dots$ consisting of piecewise linear functions, where \mathcal{S}_m ($m \geq 0$) is spanned by all Courant elements φ_θ supported on cells θ at the m -th level \mathcal{T}_m .

Utilizing these primal elements, we consider nonlinear approximation by n -term piecewise linear functions of the form $S = \sum_{j=1}^n a_{\theta_j} \varphi_{\theta_j}$, where θ_j may come from different levels and locations. Our first goal is to characterize the approximation spaces consisting of all functions with a given rate of approximation. For approximation in L_p , $p < \infty$, this is done in [11], where a collection of smoothness spaces (called B -spaces) was introduced and utilized. In this paper, we develop this theory in the more complicated case of approximation in the uniform norm ($p = \infty$). Our program consists of the following steps. First, in order to quantify the approximation process, we develop a collection of smoothness spaces $B_\tau^\alpha(\mathcal{T})$ which depend on \mathcal{T} and will govern the best approximation. Second, we prove companion Jackson and Bernstein estimates, and, third, we characterize the approximation spaces by interpolation space methods.

Our second and primary goal is, by using the B -spaces and the related techniques, to develop (or refine) algorithms for nonlinear n -term Courant element approximation so that the new algorithms are capable of achieving the rate of the best approximation. In the present paper, we develop three such algorithms for n -term Courant element approximation in L_p , which we call “threshold” ($p < \infty$), “trim and cut” ($0 < p \leq \infty$), and “push the error” ($p = \infty$) algorithms.

The first step of each of these algorithms is a decomposition step. We denote by Θ the set of all cells (supports of Courant elements) generated by \mathcal{T} . The set $(\varphi_\theta)_{\theta \in \Theta}$ is obviously redundant and, therefore, every function f has infinitely many representations of the form

$$(1.1) \quad f = \sum_{\theta \in \Theta} b_\theta(f) \varphi_\theta.$$

It is crucial to have a sufficiently efficient (sparse) initial representation of the function f that is being approximated. In our case, this means that the representation (1.1) of f should allow a realization of the corresponding B -norm $\|f\|_{B_\tau^\alpha(\mathcal{T})}$. Thus the problem of obtaining an efficient initial representation of the functions is tightly related to the development of the B -spaces. We achieve such efficiency by using good projectors into the spaces \mathcal{S}_m , $m = 0, 1, \dots$.

For completeness and comparison, we first consider the natural “threshold” algorithm for n -term Courant element approximation, which is valid only in L_p , $0 < p < \infty$. This algorithm simply takes the largest (in L_p) n -terms from (1.1).

Using the results from [11], it is easy to show that the “threshold” algorithm captures the rate of the best n -term Courant element approximation in L_p ($p < \infty$).

The second algorithm, which we call “trim and cut”, originates from the proof of the Jackson estimate in [7] and uses the following idea. First, we partition Θ through a coloring into a family of disjoint trees Θ^ν (with respect to the inclusion relation): $\Theta := \bigcup_{\nu=1}^K \Theta^\nu$. Second, we “trim” each tree by removing cells $\theta \in \Theta^\nu$ corresponding to insignificant small terms $a_\theta \varphi_\theta$ from (1.1), located near the tips of the branches. Third, we divide (“cut”) the remaining parts of each tree Θ^ν into sections of small “energy”. Finally, we rewrite the significant part of each section as a linear combination of a small number of Courant elements. The resulting terms determine the final approximant. We shall show that “trim and cut” is capable of achieving the rate of the best approximation in L_p ($0 < p \leq \infty$).

Pivotal in our development is the “push the error” algorithm, the name of which was coined by Nira Dyn. The idea for this algorithm appears in [5] and may be roughly described in L_∞ as follows. For a fixed $\varepsilon > 0$, we “push the error” with ε , starting from the coarsest level Θ_0 and proceeding to finer levels. Namely, we denote by Λ_0 the set of all $\theta \in \Theta_0$ such that $|a_\theta| > \varepsilon$ ($\|\varphi_\theta\|_\infty = 1$) and define $\mathcal{A}_0 := \sum_{\theta \in \Lambda_0} a_\theta \varphi_\theta$. Then we rewrite all remaining terms $a_\theta \varphi_\theta$ at the next level and add the resulting terms to the existing terms $a_\theta \varphi_\theta$, $\theta \in \Theta_1$. We denote the new terms by $d_\theta \varphi_\theta$, $\theta \in \Theta_1$, and select in Λ_1 all $\theta \in \Theta_1$ such that $|d_\theta| > \varepsilon$. We continue pushing the error in this way to the finer levels in the representation of f . Finally, we define our approximant by $\mathcal{A} := \sum_{m \geq 0} \mathcal{A}_j$. Thus terms $d_\theta \varphi_\theta$ with $|d_\theta| \leq \varepsilon$ are discarded only at a very fine level, and hence the error (in L_∞) is $\leq \varepsilon$.

Of course, this *naive* “push the error” algorithm cannot achieve the rate of the best approximation. However, as we shall show in §3.3 and §5, after some substantial improvements, the algorithm is capable of achieving the rate of convergence of the best n -term Courant element approximation in the uniform norm.

A focal point of our development is the characterization of the approximation spaces generated by the best n -term Courant element approximation in L_∞ and the characterization of certain approximation spaces associated with the three algorithms developed, which show that they capture the rate of convergence of the best approximation.

The outline of the paper is as follows. In §2, we collect all facts needed regarding multilevel triangulations, local approximation, quasi-interpolants, and B -spaces. In §3, we develop and explore the three algorithms for nonlinear n -term Courant element approximation: “threshold” algorithm (in §3.1), “trim and cut” algorithm (in §3.2), and “push the error” algorithm (in §3.3). Section 4 is devoted to establishing Jackson and Bernstein inequalities in order to study best n -term Courant element approximation. In §5, we show that the three algorithms capture the rate of the best n -term Courant element approximation and identify the associated approximation spaces as B -spaces. In §6, we discuss some of the main issues of nonlinear Courant element approximation. We postpone until the Appendix the proof of an important coloring lemma used in §3.2 for tree approximation in the “trim and cut” algorithm.

For convenience, we use the convention that positive constants are denoted by c, c_1, \dots throughout and they may vary at every occurrence. The notation $A \approx B$ means that $c_1 A \leq B \leq c_2 A$.

2. PRELIMINARIES

In this section we collect all the facts needed regarding multilevel triangulations, local approximation, quasi-interpolants, and other results which were developed in [11] and earlier papers. The essentials are presented for clarity but without proofs.

2.1. Triangulations. By definition $E \subset \mathbb{R}^2$ is a bounded polygonal domain if E can be represented as the union of a finite set \mathcal{T}_0 of closed triangles with disjoint interiors: $E = \bigcup_{\Delta \in \mathcal{T}_0} \Delta$. We shall always assume that there exists an initial triangulation \mathcal{T}_0 of E of this form. We call

$$\mathcal{T} = \bigcup_{m=0}^{\infty} \mathcal{T}_m$$

a *multilevel triangulation* of E with levels (\mathcal{T}_m) if the following conditions are fulfilled:

- (a) Every level \mathcal{T}_m is a partition (or triangulation) of E , that is, $E = \bigcup_{\Delta \in \mathcal{T}_m} \Delta$ and \mathcal{T}_m consists of closed triangles with disjoint interiors.
- (b) The levels (\mathcal{T}_m) of \mathcal{T} are nested, i.e., \mathcal{T}_{m+1} is a refinement of \mathcal{T}_m .
- (c) Each triangle $\Delta \in \mathcal{T}_m$ has at least two and at most M_0 children (sub-triangles) in \mathcal{T}_{m+1} , where $M_0 \geq 4$ is a constant.
- (d) The valence N_v of each vertex v of any triangle $\Delta \in \mathcal{T}_m$ (the number of the triangles from \mathcal{T}_m that share v as a vertex) is at most N_0 , where $N_0 \geq 3$ is a constant.
- (e) *No-hanging-vertices condition*: No vertex of any triangle $\Delta \in \mathcal{T}_m$ that belongs to the interior of E lies in the interior of an edge of another triangle from \mathcal{T}_m .

We denote by \mathcal{V}_m the set of all vertices of triangles from \mathcal{T}_m , where if $v \in \mathcal{V}_m$ is on the boundary of E , we include in \mathcal{V}_m as many copies of v as is its multiplicity. With this understanding, we set $\mathcal{V} = \bigcup_{m \geq 0} \mathcal{V}_m$.

We now introduce three types of multilevel nested triangulations which will play an essential role in our developments:

• **Locally regular triangulations.** We call a multilevel triangulation $\mathcal{T} = \bigcup_{m \geq 0} \mathcal{T}_m$ of E , a compact polygonal domain in \mathbb{R}^2 , a locally regular triangulation, or briefly an LR-triangulation, if \mathcal{T} satisfies the following additional conditions:

(i) There exist constants $0 < r < \rho < 1$ ($r \leq \frac{1}{4}$), such that for each $\Delta \in \mathcal{T}$ and any child Δ' of Δ that belongs to \mathcal{T} ,

$$(2.2) \quad r|\Delta| \leq |\Delta'| \leq \rho|\Delta|.$$

(ii) There exists a constant $0 < \delta \leq 1$ such that for each $\Delta', \Delta'' \in \mathcal{T}_m$ ($m \geq 0$) with a common vertex,

$$(2.3) \quad \delta \leq \frac{|\Delta'|}{|\Delta''|} \leq \frac{1}{\delta}.$$

• **Strong locally regular triangulations.** We call a multilevel triangulation $\mathcal{T} = \bigcup_{m \geq 0} \mathcal{T}_m$ of E , a compact polygonal domain in \mathbb{R}^2 , a strong locally regular triangulation, or simply an SLR-triangulation, if \mathcal{T} satisfies condition (2.2) and also the following condition (which replaces (2.3)):

(iii) *Affine transform angle condition:* There exists a constant $\beta = \beta(\mathcal{T}) > 0$ ($0 < \beta < \frac{\pi}{3}$) such that if $\Delta_0 \in \mathcal{T}_m$ ($m \geq 0$) and $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is an affine transform mapping Δ_0 one-to-one, onto an equilateral reference triangle, then for every triangle $\Delta \in \mathcal{T}_m$ with a common vertex with Δ_0 , we have

$$(2.4) \qquad \min \text{ angle}(A(\Delta)) \geq \beta$$

where $A(\Delta)$ is the image of Δ under A and is therefore also a triangle.

• **Regular triangulations.** By definition a multilevel triangulation \mathcal{T} of $E \subset \mathbb{R}^2$ is called a regular triangulation if \mathcal{T} satisfies the following condition:

(iv) There exists a constant $\beta = \beta(\mathcal{T}) > 0$ such that the minimal angle of each $\Delta \in \mathcal{T}$ is greater than or equal to β .

The remainder of this subsection makes several observations to better understand the nature of multilevel triangulations. First, it is clear that the classes of LR- and SLR-triangulations are each invariant under affine transforms. We next observe that each SLR-triangulation is an LR-triangulation, but that the converse statement does not hold. Moreover, each regular triangulation is an SLR-triangulation, but again the converse is in general false. Counterexamples are given in [11].

Each type of triangulation depends on several parameters which are not completely independent. For instance, the parameters of LR-triangulations are M_0 , N_0 , r , ρ , δ , and $\#\mathcal{T}_0$ (the cardinality of \mathcal{T}_0). We could set $M_0 = \frac{1}{r}$, $\rho = 1 - r$ and eliminate these as parameters, but this would tend to obscure the actual dependence of the estimates upon given triangulations.

We next briefly describe a simple standard procedure for constructing multilevel triangulations. We start from an initial triangulation \mathcal{T}_0 of the given compact polygonal domain $E \subset \mathbb{R}^2$. We then select a point on each edge of every triangle $\Delta \in \mathcal{T}_0$ and join them within Δ by edges to subdivide Δ into four children. The collection of all such children becomes the first generation of triangles, which we denote by \mathcal{T}_1 . We recursively refine in this way to produce succeeding generations $\mathcal{T}_2, \mathcal{T}_3, \dots$. The resulting collection $\mathcal{T} := \bigcup_{m \geq 0} \mathcal{T}_m$ is a multilevel triangulation of E .

It is important to know how the quantities $|\Delta|$, $\min \text{ angle}(\Delta)$, and $\max \text{ edge}(\Delta)$ of a triangle $\Delta \in \mathcal{T}$ may change as Δ moves away from a fixed triangle Δ^\diamond within the same level or through the nested refinements. Consider the case when \mathcal{T} is an LR-triangulation. Then conditions (i) and (ii) suggest a geometric rate of change of $|\Delta|$ (at the same level). In fact, the rate is polynomial [11]. Furthermore, if $\Delta', \Delta'' \in \mathcal{T}_m$ ($m \geq 1$) have a common vertex and are also children of some $\Delta \in \mathcal{T}_{m-1}$, then, as shown in [11], it is possible for Δ' to be equilateral (or close to such), but for Δ'' to have an uncontrollably sharp angle (see Figure 1).

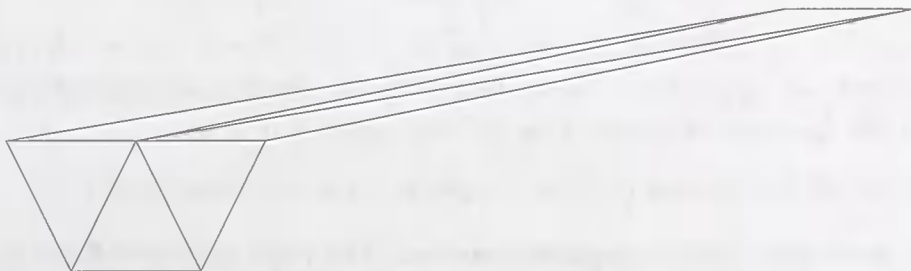


FIGURE 1. A skewed cell

If \mathcal{T} is an SLR-triangulation, the above configuration is impossible, but the triangles from \mathcal{T} still may have uncontrollably sharp angles. In this case, $\min \text{angle}(\Delta)$ changes gradually from one triangle to the adjacent ones.

For any vertex $v \in \mathcal{V}_m$ ($m \geq 0$), we denote by θ_v the cell at level m associated with v , i.e., θ_v is the union of all triangles from \mathcal{T}_m that have v as a common vertex. We denote by Θ_m the set of all such cells θ_v with $v \in \mathcal{V}_m$, and set $\Theta = \bigcup_{m \geq 0} \Theta_m$.

2.2. Local piecewise linear approximation and quasi-interpolants. We denote by Π_k the set of all algebraic polynomials of total degree less than k . We shall often refer to the following lemma (see [11]), which establishes the equivalence of different norms of polynomials over different sets.

Lemma 2.1. *Let $P \in \Pi_k$, $k \geq 1$, and $0 < p, q \leq \infty$.*

(a) *For any triangle $\Delta \subset \mathbb{R}^2$,*

$$\|P\|_{L_p(\Delta)} \approx |\Delta|^{\frac{1}{p}-\frac{1}{q}} \|P\|_{L_q(\Delta)}.$$

(b) *If Δ and Δ' are two triangles such that $\Delta' \subset \Delta$ and $|\Delta| \leq c_1 |\Delta'|$, then*

$$\|P\|_{L_p(\Delta)} \leq c \|P\|_{L_p(\Delta')}.$$

(c) *If $\Delta' \subset \Delta$ and $|\Delta'| \leq c_1 |\Delta|$ with $0 < c_1 < 1$, then*

$$\|P\|_{L_p(\Delta)} \leq c \|P\|_{L_p(\Delta \setminus \Delta')} \approx |\Delta|^{\frac{1}{p}-\frac{1}{q}} \|P\|_{L_q(\Delta \setminus \Delta')}.$$

In the above expressions, the constants depend at most on the corresponding parameters and the constant c_1 .

The no-hanging-vertices condition (e) of triangulations guarantees the existence of Courant elements. Namely, for any vertex $v \in \mathcal{V}_m$ ($m \geq 0$) there exists a unique Courant element φ_{θ_v} supported on $\theta_v \in \Theta_m$ which is the unique continuous piecewise linear function on E that is supported on θ_v and satisfies $\varphi_{\theta_v}(v) = 1$. We denote $\Phi := \Phi_{\mathcal{T}} := (\varphi_{\theta})_{\theta \in \Theta}$. We also denote by \mathcal{S}_m the space of all continuous piecewise linear functions over \mathcal{T}_m . Clearly, $S \in \mathcal{S}_m$ if and only if $S = \sum_{v \in \mathcal{V}_m} S(v) \varphi_{\theta_v}$.

Throughout the remainder of this section, we assume that \mathcal{T} is an LR-triangulation of E . We shall often use the following stability estimates for $(\varphi_{\theta})_{\theta \in \Theta_m}$.

Lemma 2.2. *Let $0 < q \leq \infty$ and $S = \sum_{\theta \in \Theta_m} a_{\theta} \varphi_{\theta}$, $m \geq 0$, with coefficients $a_{\theta} \in \mathbb{R}$.*

Then for every $\Delta \in \mathcal{T}_m$, we have

$$\|S\|_{L_q(\Delta)} \approx \left(\sum_{\theta \in \Theta_m, \theta \supset \Delta} \|a_{\theta} \varphi_{\theta}\|_q^q \right)^{\frac{1}{q}}$$

and hence

$$\|S\|_{L_q(E)} \approx \left(\sum_{\theta \in \Theta_m} \|a_{\theta} \varphi_{\theta}\|_q^q \right)^{\frac{1}{q}}$$

with constants of equivalence depending only on the parameters of \mathcal{T} . In these estimates the ℓ_q -norm is replaced by the sup-norm if $q = \infty$.

The proof of this lemma is fairly simple and can be found in [11].

• **Local piecewise linear approximation.** The local approximation by continuous piecewise linear functions will be an important tool in our further development.

For $f \in L_\eta(E)$, $\eta > 0$, and any $\Delta \in \mathcal{T}_m$ ($m \geq 0$), we denote the error of $L_\eta(\Omega_\Delta)$ -approximation to f from \mathcal{S}_m by

$$(2.5) \quad \mathbb{S}_\Delta(f)_\eta := \mathbb{S}_\Delta(f, \mathcal{T})_\eta := \inf_{S \in \mathcal{S}_m} \|f - S\|_{L_\eta(\Omega_\Delta)},$$

where Ω_Δ is the union of all triangles from \mathcal{T}_m that have a vertex in common with Δ .

• **Quasi-interpolants.** The set $\Phi_{\mathcal{T}}$ of all Courant elements is obviously redundant. To obtain a good (i.e., sparse) representation of a given function f , we shall use the following well-known quasi-interpolant:

$$(2.6) \quad Q_m(f) := Q_m(f, \mathcal{T}) := \sum_{\theta \in \Theta_m} \langle f, \tilde{\varphi}_\theta \rangle \varphi_\theta,$$

where $\langle f, g \rangle := \int_E fg$ and $(\tilde{\varphi}_\theta)$ are the duals of (φ_θ) defined by

$$(2.7) \quad \tilde{\varphi}_\theta := \sum_{\Delta \in \mathcal{T}_m, \Delta \subset \theta} \mathbb{1}_\Delta \tilde{\lambda}_{\Delta, \theta}$$

with $\tilde{\lambda}_{\Delta, \theta}$ the linear polynomial that is equal to $\frac{9}{N_v |\Delta|}$ at v_θ , the “central vertex” of θ , and equal to $-\frac{3}{N_v |\Delta|}$ at the other two vertices of Δ (recall that N_v is the valence of v). It is easily seen that

$$\langle \varphi_\theta, \tilde{\varphi}_{\theta'} \rangle = \delta_{\theta\theta'}, \text{ for } \theta, \theta' \in \Theta_m.$$

Obviously, Q_m is a linear projector, i.e., $Q_m(S) = S$ for $S \in \mathcal{S}_m$. It is crucial that $\tilde{\varphi}_\theta \in L_\infty$ and $\tilde{\varphi}_\theta$ is locally supported. Consequently, Q_m is locally bounded and provides good local approximation.

Lemma 2.3. (a) If $f \in L_\eta(E)$, $1 \leq \eta \leq \infty$, and $\Delta \in \mathcal{T}_m$, $m \geq 0$, then

$$\|Q_m(f)\|_{L_\eta(\Delta)} \leq c \|f\|_{L_\eta(\Omega_\Delta)}.$$

(b) If $0 < \eta \leq \infty$ and $g = \sum_{\Delta \in \mathcal{T}_m} \mathbb{1}_\Delta \cdot P_\Delta$ with $P_\Delta \in \Pi_2$ and $m \geq 0$, then

$$\|Q_m(g)\|_{L_\eta(\Delta)} \leq c \|g\|_{L_\eta(\Omega_\Delta)}, \text{ for } \Delta \in \mathcal{T}_m.$$

The constants above depend only on η and the parameters of \mathcal{T} .

For a proof of this lemma, see [11].

From the above lemma, we see that $Q_m : L_\eta(E) \rightarrow \mathcal{S}_m$ ($1 \leq \eta \leq \infty$) is a locally bounded linear projector. There is a well-known scheme for extending Q_m to a nonlinear projector $Q_m : L_\eta(E) \rightarrow \mathcal{S}_m$ for $0 < \eta < 1$. This is needed for nonlinear approximation in L_p ($0 < p \leq 1$). To describe this extension, let $P_{\Delta, \eta} : L_\eta(\Delta) \rightarrow \Pi_2$ ($0 < \eta \leq \infty$) be a projector (linear if $\eta \geq 1$ and nonlinear if $0 < \eta < 1$) such that

$$\|f - P_{\Delta, \eta}(f)\|_{L_\eta(\Delta)} \leq c E_2(f, \Delta) \quad \text{for } f \in L_\eta(\Delta),$$

where $E_2(f, \Delta)$ is the error of the best $L_\eta(\Delta)$ -approximation to f from Π_2 (the linear polynomials). We define

$$p_{m, \eta}(f) := \sum_{\Delta \in \mathcal{T}_m} \mathbb{1}_\Delta \cdot P_{\Delta, \eta}(f)$$

and set

$$(2.8) \quad T_{m, \eta}(f) := Q_m(p_{m, \eta}(f)), \quad \text{for } f \in L_\eta(E).$$

Clearly, $T_{m,\eta} : L_\eta(E) \rightarrow \mathcal{S}_m$ is a projector (linear if $\eta \geq 1$ and nonlinear if $0 < \eta < 1$).

The next lemma, established in [11], shows that Q_m and $T_{m,\eta}$ provide good local approximations from \mathcal{S}_m .

Lemma 2.4. (a) If $f \in L_\eta(E)$, $1 \leq \eta \leq \infty$, and $\Delta \in \mathcal{T}_m$, $m \geq 0$, then

$$\|f - Q_m(f)\|_{L_\eta(\Delta)} \leq c\mathbb{S}_\Delta(f)_\eta.$$

(b) If $f \in L_\eta(E)$, $0 < \eta \leq \infty$, and $\Delta \in \mathcal{T}_m$, $m \geq 0$, then

$$\|f - T_m(f)\|_{L_\eta(\Delta)} \leq c\mathbb{S}_\Delta(f)_\eta.$$

The constants above depend only on η and the parameters of \mathcal{T} .

The needed convergence of $Q_m(f)$ and $T_m(f)$ to f is provided by the following result (see Lemma 2.15 from [11]).

Lemma 2.5. If $f \in L_\eta(E)$, then

$$\begin{aligned} \|f - Q_m(f)\|_{L_\eta(E)} &\rightarrow 0 \text{ as } m \rightarrow \infty, & \text{if } 1 \leq \eta \leq \infty, \\ \|f - T_{m,\eta}(f)\|_{L_\eta(E)} &\rightarrow 0 \text{ as } m \rightarrow \infty, & \text{if } 0 < \eta \leq \infty. \end{aligned}$$

Now, we apply a well-known scheme for obtaining sparse Courant element representation of functions. We define

$$(2.9) \quad q_m := Q_m - Q_{m-1} \text{ and } t_{m,\eta} := T_{m,\eta} - T_{m-1,\eta}, \text{ for } m \geq 0,$$

where $Q_{-1} := 0$ and $T_{-1,\eta} := 0$. Clearly, $q_m(f), t_{m,\eta}(f) \in \mathcal{S}_m$.

For a given function $f \in L_\eta(E)$, $1 \leq \eta \leq \infty$, we define the sequence $\mathbf{b}(f) := (b_\theta(f))_{\theta \in \Theta_m}$ from the expression

$$(2.10) \quad q_m(f) =: \sum_{\theta \in \Theta_m} b_\theta(f) \varphi_\theta, \quad m \geq 0.$$

Using Lemma 2.5, we have

$$(2.11) \quad f = \sum_{m \geq 0} q_m(f) = \sum_{m \geq 0} \sum_{\theta \in \Theta_m} b_\theta(f) \varphi_\theta \quad \text{in } L_\eta.$$

If $f \in L_\eta(E)$, $0 < \eta < 1$, we define the sequence $\mathbf{b}_\eta(f) := (b_{\theta,\eta}(f))_{\theta \in \Theta_m}$ by

$$(2.12) \quad t_{m,\eta}(f) =: \sum_{\theta \in \Theta_m} b_{\theta,\eta}(f) \varphi_\theta, \quad m \geq 0,$$

and again by Lemma 2.5, we have

$$(2.13) \quad f = \sum_{m \geq 0} t_{m,\eta}(f) = \sum_{m \geq 0} \sum_{\theta \in \Theta_m} b_{\theta,\eta}(f) \varphi_\theta \quad \text{in } L_\eta.$$

Clearly, $\mathbf{b}(\cdot)$ is a linear operator while $\mathbf{b}_\eta(\cdot)$ ($0 < \eta < 1$) is nonlinear.

2.3. B -spaces. In this section, we include the necessary tools for the B -spaces which we need for nonlinear n -term Courant element approximation. The B -spaces over multilevel nested triangulations of \mathbb{R}^2 are introduced in [11] and used there for nonlinear n -term Courant element approximation in $L_p(\mathbb{R}^2)$ ($0 < p < \infty$). In the present paper, we shall use the B -spaces for n -term Courant element approximation in $L_p(E)$ ($0 < p \leq \infty$), where E is a compact polygonal domain in \mathbb{R}^2 . We shall put the emphasis on approximation in the uniform norm ($p = \infty$). There are three types of B -spaces (skinny, slim, and fat B -spaces) that were introduced in [11] to serve different purposes. For Courant element approximation, we need the slim B -spaces, which we shall simply call B -spaces.

Throughout this paper, we assume that \mathcal{T} is an LR -triangulation of a compact polygonal domain E in \mathbb{R}^2 . Moreover, the B -spaces $B_\tau^\alpha(\mathcal{T})$, with parameter set $1/\tau := \alpha + 1/p$ according to two specific choices: (a) $p = \infty$ and $\alpha \geq 1$; or (b) $0 < p < \infty$ and $\alpha > 0$, will arise naturally in our algorithms and error estimates. These spaces have several equivalent definitions, which we briefly describe.

• **Definition of $B_\tau^\alpha(\mathcal{T})$ via local approximation.** We define $B_\tau^\alpha(\mathcal{T})$ as the set of all functions $f \in L_\tau(E)$ such that

$$(2.14) \qquad |f|_{B_\tau^\alpha(\mathcal{T})} := \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{-\alpha} \mathbb{S}_\Delta(f)_\tau)^\tau \right)^{1/\tau} < \infty,$$

where $\mathbb{S}_\Delta(f)_\tau$ is the error of $L_\tau(\Omega_\Delta)$ -approximation (local) to f from \mathcal{S}_m for $\Delta \in \mathcal{T}_m$ (see (2.5)). It is readily seen that $|f + g|_{B_\tau^\alpha}^{\tau^*} \leq |f|_{B_\tau^\alpha}^{\tau^*} + |g|_{B_\tau^\alpha}^{\tau^*}$ with $\tau^* := \min\{\tau, 1\}$, and $|f + s|_{B_\tau^\alpha} = |f|_{B_\tau^\alpha}$ for $s \in \mathcal{S}_0$. Hence $|\cdot|_{B_\tau^\alpha}$ is a semi-norm if $\tau \geq 1$ and a semi-quasi-norm if $\tau < 1$.

By Theorems 2.7 and 2.9 below, it follows that if $f \in B_\tau^\alpha(\mathcal{T})$, then $f \in L_p(E)$. Therefore, it is natural to define a (quasi-)norm in $B_\tau^\alpha(\mathcal{T})$ by

$$(2.15) \qquad \|f\|_{B_\tau^\alpha(\mathcal{T})} := \|f\|_p + |f|_{B_\tau^\alpha(\mathcal{T})}.$$

More generally, for $0 < \eta < p$, we define

$$(2.16) \qquad N_{\mathbb{S},\eta}(f, \mathcal{T}) := \|f\|_p + \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{1/p-1/\eta} \mathbb{S}_\Delta(f)_\eta)^\tau \right)^{1/\tau}.$$

Evidently, $N_{\mathbb{S},\tau}(f, \mathcal{T}) = \|f\|_{B_\tau^\alpha(\mathcal{T})}$. When clear from the context, we use $N_{\mathbb{S},\tau}(f)$.

• **Definition of norm in $B_\tau^\alpha(\mathcal{T})$ via atomic decomposition.** For $f \in L_\tau(E)$, we define

$$(2.17) \qquad N_\Phi(f) := \inf_{f = \sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{\theta \in \Theta} (|\theta|^{-\alpha} \|c_\theta \varphi_\theta\|_\tau)^\tau \right)^{1/\tau},$$

where the infimum is taken over all representations $f = \sum_{\theta \in \Theta} c_\theta \varphi_\theta$ in $L_\tau(E)$. Note that the existence of such representations of f follows by (2.11) and (2.13). By Theorem 2.7 below,

$$\sum_{\theta \in \Theta} (|\theta|^{-\alpha} \|c_\theta \varphi_\theta\|_\tau)^\tau < \infty \quad \text{implies} \quad \left\| \sum_{\theta \in \Theta} |c_\theta \varphi_\theta(\cdot)| \right\|_p < \infty,$$

and hence $f \in L_p(E)$ and the series $\sum_{\theta \in \Theta} |c_\theta \varphi_\theta(\cdot)|$ converges a.e. and in $L_p(E)$. Therefore, the way in which the terms of the series are ordered is not essential,

and the convergence in $L_\tau(E)$ implies a stronger (absolute) convergence in $L_p(E)$ ($\tau < p$). By Lemma 2.1, it follows that

$$\begin{aligned} N_\Phi(f) &\approx \inf_{f=\sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{\theta \in \Theta} (|\theta|^{1/p} |c_\theta|)^\tau \right)^{1/\tau} \\ (2.18) \quad &\approx \inf_{f=\sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{\theta \in \Theta} \|c_\theta \varphi_\theta\|_p^\tau \right)^{1/\tau}. \end{aligned}$$

If $p = \infty$, then

$$N_\Phi(f) \approx \inf_{f=\sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{\theta \in \Theta} |c_\theta|^\tau \right)^{1/\tau}.$$

• **Definition of norms in $B_\tau^\alpha(\mathcal{T})$ via projectors.** For $f \in L_\eta(E)$, we let

$$(2.19) \quad f = \sum_{\theta \in \Theta} b_{\theta, \eta}(f) \varphi_\theta$$

be the representation of f from (2.11) if $\eta \geq 1$ and from (2.13) if $0 < \eta < 1$. We define

$$(2.20) \quad N_{Q, \tau}(f) := \left(\sum_{\theta \in \Theta} (|\theta|^{-\alpha} \|b_{\theta, \tau}(f) \varphi_\theta\|_\tau)^\tau \right)^{1/\tau}$$

and, more generally (in accordance with (2.16)),

$$(2.21) \quad N_{Q, \eta}(f) := \left(\sum_{\theta \in \Theta} (|\theta|^{1/p-1/\eta} \|b_{\theta, \eta}(f) \varphi_\theta\|_\eta)^\tau \right)^{1/\tau}.$$

By Lemmas 2.1 and 2.2, we have

$$(2.22) \quad N_{Q, \eta}(f) \approx \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{1/p-1/\eta} \|q_\Delta(f)\|_{L_\eta(\Delta)})^\tau \right)^{1/\tau}, \quad \text{if } \eta \geq 1,$$

$$(2.23) \quad N_{Q, \eta}(f) \approx \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{1/p-1/\eta} \|t_{\Delta, \eta}(f)\|_{L_\eta(\Delta)})^\tau \right)^{1/\tau}, \quad \text{if } 0 < \eta < 1,$$

and

$$(2.24) \quad N_{Q, \eta}(f) \approx \left(\sum_{\theta \in \Theta} (|\theta|^{1/p} |b_{\theta, \eta}(f)|)^\tau \right)^{1/\tau} \approx \left(\sum_{\theta \in \Theta} \|b_{\theta, \eta}(f) \varphi_\theta\|_p^\tau \right)^{1/\tau}.$$

In the most interesting case of $p = \infty$,

$$(2.25) \quad N_{Q, \eta}(f) \approx \left(\sum_{\theta \in \Theta} |b_{\theta, \eta}(f)|^\tau \right)^{1/\tau}.$$

• **General B -spaces.** A more general B -space $B_{pq}^\alpha(\mathcal{T})$, $\alpha > 0$, $0 < p, q \leq \infty$, is defined as the set of all $f \in L_p(E)$ such that

$$\|f\|_{B_{pq}^\alpha(\mathcal{T})} := \inf_{f=\sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{m \in \mathbb{Z}} \left[2^{m\alpha} \left(\sum_{\theta \in \Theta, 2^{-m} \leq |\theta| < 2^{-m+1}} \|c_\theta \varphi_\theta\|_p^p \right)^{1/p} \right]^q \right)^{1/q} < \infty,$$

where the ℓ_q -norm is replaced by the sup-norm if $q = \infty$. In this paper, we do not need the B -spaces in such generality.

• **Embedding theorems and equivalence of norms.** We recall our assumptions. We have $0 < p \leq \infty$, and $\alpha \geq 1$ if $p = \infty$ and $\alpha > 0$ if $p < \infty$. In both cases, $1/\tau := \alpha + 1/p$ ($1/\tau := \alpha$ if $p = \infty$). We record estimates and embeddings from [11], along with the necessary modifications, which are necessary for the development of the main results of this paper. The first embedding result appears as Theorem 2.16 in [11].

Theorem 2.6. *For $0 < \tau < p$ or $p = \infty, \tau \leq 1$, then for any sequence of real numbers $(c_\theta)_{\theta \in \Theta}$, we have*

$$(2.26) \quad \left\| \sum_{\theta \in \Theta} |c_\theta| \varphi_\theta \right\|_p \leq c \left(\sum_{\theta \in \Theta} \|c_\theta \varphi_\theta\|_p^\tau \right)^{1/\tau},$$

where c depends only on τ , p , and the parameters of \mathcal{T} .

Theorem 2.7. *If $f \in L_\eta(E)$ with $0 < \eta < p$, and $N_{Q,\eta}(f) < \infty$, then $f \in L_p(E)$ ($f \in C(E)$ if $p = \infty$), and f has the representation $f = \sum_{\theta \in \Theta} b_{\theta,\eta}(f) \varphi_\theta$ with the series converging absolutely a.e. in E and in L_p (respectively, in $C(E)$), and*

$$(2.27) \quad \|f\|_p \leq \left\| \sum_{\theta \in \Theta} |b_{\theta,\eta}(f)| \varphi_\theta \right\|_p \leq c N_{Q,\eta}(f),$$

where c is independent of f .

Proof. For $0 < p < \infty$, the result follows from (2.11), (2.13), and Theorem 2.9 below. If $p = \infty$, the theorem follows by (2.11), (2.13), (2.25), and the following estimates:

$$\left\| \sum_{\theta \in \Theta} b_{\theta,\eta}(f) \varphi_\theta \right\|_\infty \leq \left(\sum_{\theta \in \Theta} |b_{\theta,\eta}(f)|^\tau \right)^{1/\tau} \leq c N_{Q,\eta}(f) \quad (\tau = 1/\alpha \leq 1). \quad \square$$

Remark 2.8. It is easily seen that Theorem 2.7 is not true when $p = \infty$ and $\alpha < 1$. For this reason we impose the restriction $\alpha \geq 1$ when $p = \infty$ throughout.

Theorem 2.9. *The norms $\|\cdot\|_{B_\tau^\alpha(\mathcal{T})}$, $N_{\mathbb{S},\eta}(\cdot)$ ($0 < \eta < p$), $N_\Phi(\cdot)$, and $N_{Q,\eta}(\cdot)$ ($0 < \eta < p$), defined in (2.15), (2.16), (2.17), and (2.21), are equivalent with constants of equivalence depending only on p , τ , η , and the parameters of \mathcal{T} .*

Proof. One proceeds exactly as in [11] (see the proof of Theorem 2.17 of that reference) and proves that

$$(2.28) \quad \begin{aligned} |f|_{B_\tau^\alpha(\mathcal{T})} &\approx \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{1/p-1/\eta} \mathbb{S}_\Delta(f)_\eta)^\tau \right)^{1/\tau} \approx \inf_{f = \sum_{\theta \in \Theta} c_\theta \varphi_\theta} \left(\sum_{\theta \in \Theta \setminus \Theta_0} \|c_\theta \varphi_\theta\|_p^\tau \right)^{1/\tau} \\ &\approx \left(\sum_{\theta \in \Theta \setminus \Theta_0} \|b_{\theta,\eta}(f) \varphi_\theta\|_p^\tau \right)^{1/\tau}, \end{aligned}$$

provided $0 < \eta < p$. To obtain the norm estimate from these semi-norm equivalences, we use Theorem 2.6 to give $\|f\|_p \leq c N_\Phi(f)$. Using this, (2.28), and the remark after the definition of $N_\Phi(f)$ in (2.17), we obtain

$$\|f\|_{B_\tau^\alpha(\mathcal{T})} \approx N_{\mathbb{S},\eta}(f) \leq c N_\Phi(f) \leq c N_{Q,\eta}(f), \quad 0 < \eta < p.$$

For the reverse inequality, we use Lemma 2.2, Theorem 2.6, and (2.28) to obtain

$$\begin{aligned} \left(\sum_{\theta \in \Theta_0} \|b_{\theta, \eta}(f) \varphi_{\theta}\|_p^{\tau} \right)^{1/\tau} &\leq c(\#\Theta_0, \tau, p) \left(\sum_{\theta \in \Theta_0} \|b_{\theta, \eta}(f) \varphi_{\theta}\|_p^p \right)^{1/p} \\ &\leq c \left\| \sum_{\theta \in \Theta_0} b_{\theta, \eta}(f) \varphi_{\theta} \right\|_p \leq c \left(\|f\|_p + \left\| \sum_{\theta \in \Theta \setminus \Theta_0} b_{\theta, \eta}(f) \varphi_{\theta} \right\|_p \right) \\ &\leq c \|f\|_p + c \left(\sum_{\theta \in \Theta \setminus \Theta_0} \|b_{\theta, \eta}(f) \varphi_{\theta}\|_p^{\tau} \right)^{1/\tau} \\ &\leq c \|f\|_{B_{\tau}^{\alpha}(\mathcal{T})}. \end{aligned}$$

This and (2.28) imply $N_{Q, \eta}(f) \leq c \|f\|_{B_{\tau}^{\alpha}(\mathcal{T})}$. \square

The next embedding theorem of Sobolev type follows immediately from (2.18) or (2.24).

Theorem 2.10. *For $0 < \alpha_0 < \alpha_1$ and $\tau_j := (\alpha_j + 1/p)^{-1}$, $j = 0, 1$, we have the continuous embedding*

$$(2.29) \quad B_{\tau_1}^{\alpha_1}(\mathcal{T}) \subset B_{\tau_0}^{\alpha_0}(\mathcal{T}),$$

i.e., if $f \in B_{\tau_1}^{\alpha_1}(\mathcal{T})$, then $f \in B_{\tau_0}^{\alpha_0}(\mathcal{T})$ and $\|f\|_{B_{\tau_0}^{\alpha_0}(\mathcal{T})} \leq c \|f\|_{B_{\tau_1}^{\alpha_1}(\mathcal{T})}$.

• **Interpolation.** We first recall some basic definitions from the real interpolation method. We refer the reader to [2] and [1] as general references for interpolation theory. For a pair of quasi-normed spaces X_0, X_1 , embedded in a Hausdorff space, the space $X_0 + X_1$ is defined as the collection of all functions f that can be represented as $f_0 + f_1$ with $f_0 \in X_0$ and $f_1 \in X_1$. The quasi-norm in $X_0 + X_1$ is defined by

$$\|f\|_{X_0 + X_1} := \|f\|_{X_0 + X_1} + \inf_{f=f_0+f_1} \|f_0\|_{X_0} + \|f_1\|_{X_1}.$$

The K -functional is defined for each $f \in X_0 + X_1$ and $t > 0$ by

$$(2.30) \quad K(f, t) := K(f, t; X_0, X_1) := \inf_{f=f_0+f_1} \|f_0\|_{X_0} + t \|f_1\|_{X_1}.$$

The real interpolation space $(X_0, X_1)_{\lambda, q}$ with $0 < \lambda < 1$ and $0 < q \leq \infty$ is defined as the set of all $f \in X_0 + X_1$ such that

$$\|f\|_{(X_0, X_1)_{\lambda, q}} := \left(\int_0^{\infty} (t^{-\lambda} K(f, t))^q \frac{dt}{t} \right)^{1/q} < \infty,$$

where the L_q -norm is replaced by the sup-norm if $q = \infty$.

It is easily seen that if $X_1 \subset X_0$ (X_1 continuously embedded in X_0), then $K(f, t) \approx \|f\|_{X_0}$ for $f \in X_0$ and $t \geq 1$, and, consequently,

$$(2.31) \quad \|f\|_{(X_0, X_1)_{\lambda, q}} \approx \|f\|_{X_0} + \left(\sum_{\nu=0}^{\infty} [2^{\nu\lambda} K(f, 2^{-\nu})]^q \right)^{1/q}.$$

Theorem 2.11. *Suppose $0 < p \leq \infty$ and further assume that both $\alpha_0, \alpha_1 \geq 1$ in the case $p = \infty$, and $\alpha_0, \alpha_1 > 0$ otherwise. Furthermore, let $\tau_j := (\alpha_j + 1/p)^{-1}$, $j = 0, 1$. Then*

$$(2.32) \quad (B_{\tau_0}^{\alpha_0}(\mathcal{T}), B_{\tau_1}^{\alpha_1}(\mathcal{T}))_{\lambda, \tau} = B_{\tau}^{\alpha}(\mathcal{T})$$

with equivalent norms, provided $\alpha = (1 - \lambda)\alpha_0 + \lambda\alpha_1$ with $0 < \lambda < 1$ and $\tau := (\alpha + 1/p)^{-1}$.

Proof. We shall prove (2.32) only in the case $p > 1$. For a proof of (2.32) when $p \leq 1$, see [3].

We shall use the abbreviated notation $B^\alpha := B^\alpha_\tau(\mathcal{T})$ and $B^{\alpha_j} := B^{\alpha_j}_{\tau_j}(\mathcal{T})$, $j = 0, 1$. Also, we denote by ℓ_q the space of all sequences $\mathbf{a} = (a_\theta)_{\theta \in \Theta}$ of real numbers such that $\|\mathbf{a}\|_{\ell_q} := (\sum_{\theta \in \Theta} |a_\theta|^q)^{1/q} < \infty$.

We set $\eta := 1$ and normalize the Courant elements in L_p , that is, $\|\varphi_\theta\|_p = 1$. We also renormalize the duals $\tilde{\varphi}_\theta$ from (2.7) accordingly. We denote again by $\mathbf{b}(f) = (b_\theta)_{\theta \in \Theta}$ the sequence from (2.10) with respect to the normalized Courant elements. By (2.24), Theorem 2.7, and Theorem 2.9, if $f \in B^{\alpha_j}$, $j = 0, 1$, then

$$(2.33) \quad f = \sum_{\theta \in \Theta} b_\theta(f) \varphi_\theta \quad \text{and} \quad \|f\|_{B^{\alpha_j}} \approx \|\mathbf{b}(f)\|_{\ell_{\tau_j}},$$

recalling that the elements φ_θ are normalized in L_p . The corresponding statement holds for functions $f \in B^\alpha$ as well.

We shall next employ the following interpolation theorem (see, e.g., §5.1 of [1] or [2]) which follows directly from the definition of the K -functional and the norms of the interpolation spaces. Suppose T is a linear operator which boundedly maps X_0 into Y_0 and X_1 into Y_1 , where (X_0, X_1) and (Y_0, Y_1) are couples of quasi-normed spaces as above. Then for $0 < \lambda < 1$ and $0 < q \leq \infty$, T boundedly maps $(X_0, X_1)_{\lambda, q}$ into $(Y_0, Y_1)_{\lambda, q}$.

We introduce linear operators \mathcal{I} and \mathcal{P} as follows: \mathcal{I} is defined by $\mathcal{I}(f)_\theta := b_\theta(f)$, $\theta \in \Theta$, and \mathcal{P} is given by $\mathcal{P}(\mathbf{a}) := \sum_{\theta \in \Theta} a_\theta \varphi_\theta$, $\mathbf{a} = (a_\theta)_{\theta \in \Theta}$. By (2.33), $\|\mathbf{b}(f)\|_{\ell_{\tau_j}} \leq c\|f\|_{B^{\alpha_j}}$ for $f \in B^{\alpha_j}$, $j = 0, 1$, and hence $\mathcal{I} : B^{\alpha_j} \rightarrow \ell_{\tau_j}$ (boundedly). By the above-mentioned interpolation theorem,

$$(2.34) \quad \mathcal{I} : (B^{\alpha_0}, B^{\alpha_1})_{\lambda, \tau} \rightarrow (\ell_{\tau_0}, \ell_{\tau_1})_{\lambda, \tau} \quad (\text{boundedly}).$$

Similarly, if $\mathbf{a} \in \ell_{\tau_j}$, then by Theorems 2.7 and 2.9, we may conclude that $\mathcal{P}(\mathbf{a}) \stackrel{L_p}{=} \sum_{\theta \in \Theta} a_\theta \varphi_\theta$ is well defined. So if we set $f = \mathcal{P}(\mathbf{a})$, then

$$\|\mathcal{P}(\mathbf{a})\|_{B^{\alpha_j}} \leq c \inf_{f = \sum_{\theta \in \Theta} c_\theta \varphi_\theta} \|(c_\theta)_{\theta \in \Theta}\|_{\ell_{\tau_j}} \leq c \|\mathbf{a}\|_{\ell_{\tau_j}}, \quad j = 0, 1.$$

Thus $\mathcal{P} : \ell_{\tau_j} \rightarrow B^{\alpha_j}$ (boundedly), and by interpolation

$$(2.35) \quad \mathcal{P} : (\ell_{\tau_0}, \ell_{\tau_1})_{\lambda, \tau} \rightarrow (B^{\alpha_0}, B^{\alpha_1})_{\lambda, \tau} \quad (\text{boundedly}).$$

Finally, we recall the well-known interpolation result (see, e.g., [2], [1]):

$$(2.36) \quad (\ell_{\tau_0}, \ell_{\tau_1})_{\lambda, \tau} = \ell_\tau, \quad \text{where } \frac{1}{\tau} = \frac{1-\lambda}{\tau_0} + \frac{\lambda}{\tau_1} \text{ with } 0 < \lambda < 1.$$

Clearly, (2.32) follows by (2.33)-(2.36). □

• **Skinny B -spaces.** The skinny B -spaces were introduced in [11] and used for characterization of nonlinear (discontinuous) piecewise polynomial approximation on \mathbb{R}^2 . We next adapt that definition to the case of approximation on a compact polygonal domain $E \subset \mathbb{R}^2$. Suppose \mathcal{T} is a multilevel nested triangulation of E which additionally satisfies condition (2.2) (see §2.1 and [11]). The skinny B -space

$\mathcal{B}_\tau^{\alpha k}(\mathcal{T})$, where $k \geq 1$ and α and τ are as above, is defined as the set of all $f \in L_\tau(E)$ such that

$$(2.37) \qquad |f|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} := \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{-\alpha} \omega_k(f, \Delta)_\tau)^\tau \right)^{1/\tau} < \infty,$$

where $\omega_k(f, \Delta)_\tau$ is a k th modulus of smoothness of f in $L_\tau(\Delta)$, defined by

$$\omega_k(f, \Delta)_\tau := \sup_{h \in \mathbb{R}^2} \|\Delta_h^k(f, \cdot)\|_{L_\tau(\Delta)}$$

and $\Delta_h^k(f, \cdot)$ is the k th difference of f . The norm in $\mathcal{B}_\tau^{\alpha k}(\mathcal{T})$ is defined by $\|\cdot\|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})} := \|\cdot\|_p + |\cdot|_{\mathcal{B}_\tau^{\alpha k}(\mathcal{T})}$.

• **Fat B -spaces: The link to Besov spaces.** Suppose \mathcal{T} is an SLR-triangulation of a compact polygonal domain $E \subset \mathbb{R}^2$. Similarly as in [11], we define the fat B -space $\mathbb{B}_\tau^{\alpha k}(\mathcal{T})$, where $k \geq 1$ and α and τ are as above, as the set of all functions $f \in L_\tau(E)$ such that

$$(2.38) \qquad |f|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})} := \left(\sum_{\Delta \in \mathcal{T}} (|\Delta|^{-\alpha} \omega_k(f, \Omega_\Delta)_\tau)^\tau \right)^{1/\tau} < \infty.$$

We endow $\mathbb{B}_\tau^{\alpha k}(\mathcal{T})$ with the norm $\|\cdot\|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})} := \|\cdot\|_p + |\cdot|_{\mathbb{B}_\tau^{\alpha k}(\mathcal{T})}$. Using Whitney’s theorem, it readily follows that $c_1 \omega_2(f, \Delta)_\tau \leq \mathbb{S}_\Delta(f)_\tau \leq c_2 \omega_2(f, \Omega_\Delta)_\tau$, and hence $|f|_{\mathcal{B}_\tau^{\alpha 2}(\mathcal{T})} \leq c |f|_{\mathbb{B}_\tau^{\alpha 2}(\mathcal{T})} \leq c |f|_{\mathcal{B}_\tau^{\alpha 2}(\mathcal{T})}$. The space $\mathbb{B}_\tau^{\alpha 2}(\mathcal{T})$ is a natural candidate to replace $B_\tau^\alpha(\mathcal{T})$ in nonlinear n -term Courant element approximation. This is, however, only possible for sufficiently small α ($0 < \alpha < \alpha_0$). Otherwise $\mathbb{B}_\tau^{\alpha 2}(\mathcal{T})$ is too “fat” and cannot do the job. Finally, we note that if \mathcal{T} is a regular triangulation and $0 < \alpha < k$, then $\mathbb{B}_\tau^{\alpha k}(\mathcal{T})$ coincides with the Besov space $B_\tau^{2\alpha}(L_\tau)$. For a more complete discussion of this and other related issues, see [11].

3. ALGORITHMS FOR n -TERM COURANT APPROXIMATION

• **Decomposition step for all approximation algorithms.** The first step of each of the three approximation algorithms that we consider in this section is a decomposition step. This step is not trivial, since the set $\Phi_\mathcal{T} := (\varphi_\theta)_{\theta \in \Theta}$ of all Courant elements is redundant and, therefore, each function has infinitely many representations using Courant elements. For each algorithm, it is crucial to have a sufficiently efficient initial representation of the function f that is being approximated. This means that the representation of f should allow a realization of the corresponding B -norm.

To construct the initial representation, we consider two cases of metric approximation. If the approximation takes place in L_p , $1 < p \leq \infty$, we utilize the decomposition of f via quasi-interpolation from (2.11) with $1 \leq \eta < p$, while if $0 < p \leq 1$, we use (2.13) with $0 < \eta < p$. In both cases, we have an initial desirable sparse representation of f of the form

$$(3.1) \qquad f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta, \quad b_\theta = b_\theta(f),$$

which allows a realization of the B -norm (see (2.24)–(2.25), and Theorem 2.9). For the remainder of this section, in order to more easily track the dependency of the

constants appearing in the inequalities, we redefine $\|f\|_{B_\tau^\alpha(\mathcal{T})}$ by

$$(3.2) \quad \|f\|_{B_\tau^\alpha(\mathcal{T})} := \left(\sum_{\theta \in \Theta} (|b_\theta| |\theta|^{1/p})^\tau \right)^{1/\tau} \approx \left(\sum_{\theta \in \Theta} \|b_\theta \varphi_\theta\|_p^\tau \right)^{1/\tau},$$

which is an equivalent norm in $B_\tau^\alpha(\mathcal{T})$ (see Theorem 2.9). Without loss of generality, we may assume (when needed) that there is a final level Θ_L ($L < \infty$) in (3.1).

3.1. “Threshold” algorithm ($p < \infty$ only). In this algorithm we utilize the usual thresholding strategy used for n -term approximation from a basis in L_p ($1 < p < \infty$). The resulting procedure performs extremely well, due to the sparse representation realized by the first step. We note, however, that the derived error estimates involve constants that depend on p and become unbounded as $p \rightarrow \infty$. The “push the error” and “trim and cut” algorithms described later in this section will be shown to achieve the corresponding estimates for the uniform norm ($p = \infty$). For this subsection we therefore assume that $f \in L_p$, $0 < p < \infty$.

• **Description of the “threshold” algorithm.**

Step 1. (*Decompose*) We use the decomposition of $f \in L_p(E)$ from (3.1).

Step 2. (*Select the n largest terms*) We order the terms $(b_\theta \varphi_\theta)_{\theta \in \Theta}$ in a sequence $(b_{\theta_j} \varphi_{\theta_j})_{j=1}^\infty$ so that

$$(3.3) \quad \|b_{\theta_1} \varphi_{\theta_1}\|_p \geq \|b_{\theta_2} \varphi_{\theta_2}\|_p \geq \cdots.$$

Then we define the approximant $A_n^T(f)_p$ by $A_n^T(f)_p := \sum_{j=1}^n b_{\theta_j} \varphi_{\theta_j}$.

• **Error estimation for the “threshold” algorithm.** We denote the corresponding error of approximation of this threshold algorithm by

$$\mathbb{A}_n^T(f)_p := \|f - A_n^T(f)_p\|_p.$$

The argument used in establishing the Jackson error estimate in [11] may be modified in obvious ways to prove the following error estimate.

Theorem 3.1. *If $f \in B_\tau^\alpha(\mathcal{T})$, $\alpha > 0$, $1/\tau := \alpha + 1/p$ ($0 < p < \infty$), then*

$$(3.4) \quad \mathbb{A}_n^T(f)_p \leq cn^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})},$$

where c depends on α , p , and the parameters of \mathcal{T} .

In §5, we shall need the following result:

Lemma 3.2. *If $f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta$ is the decomposition of f from (3.1), then*

$$\mathbb{A}_{2n}^T(f)_p \leq cn^{-\alpha} \left(\sum_{j=n+1}^\infty \|b_{\theta_j} \varphi_{\theta_j}\|_p^\tau \right)^{1/\tau},$$

where $(b_{\theta_j} \varphi_{\theta_j})_{j=1}^\infty$ is as in Step 2 and c depends on α , p , and the parameters of \mathcal{T} .

Proof. Applying Theorem 3.4 from [11] to $(b_{\theta_j} \varphi_{\theta_j})_{j=n+1}^\infty$ immediately provides the desired result. \square

Remark 3.3. As we have mentioned, the main drawback of the “threshold” algorithm is that it is not applicable to approximation in the uniform norm, since the constant $c = c(\alpha, p)$ in (3.4) tends to infinity as $p \rightarrow \infty$ and the performance of the algorithm deteriorates as p gets large. The obvious reason for this behavior is that

f can be built out of many terms ($b_\theta \varphi_\theta$) which have small coefficients and are supported at the same location. These terms can pile up to an essential contribution, but the algorithm will fail to anticipate their future significance.

3.2. “Trim and cut (the tree)” algorithm. The idea of this algorithm has its origins in the proof of the Jackson estimate in [7] (see §5, pages 272-276). The approximation considered there is by wavelets or splines over a uniform partition in the uniform norm. We shall refine this idea to develop an algorithm for n -term Courant element approximation in $L_p(E)$, $0 < p \leq \infty$, over LR-triangulations. We begin with a brief description of the algorithm and then elaborate on the details of each of the main steps.

• **Description of the “trim and cut” algorithm.**

- Step 1.** (*Decompose*). We use the common decomposition of $f \in L_p(E)$ given in (3.1).
- Step 2.** (*Organize the cells of Θ into manageable trees Θ^ν*). We develop an algorithm (procedure) for coloring the cells of Θ in such a way that the cells of the same color form a tree structure as described in Lemma 3.4 below. This organization greatly simplifies the management of the estimates, both the approximation construction and the enumeration of “active” Courant elements in our approximant.
- Step 3.** (*Trim each tree*). Since all the elements may initially affect the B -space norm of a function, we need to preprocess each tree by pruning all branches which may have many leaves, but do not make a significant contribution to the norm of the function f . We do this by running a stopping time argument from the finest level to a coarser level, until a significant cumulative contribution is met. We prune the branch just below that element.
- Step 4.** (*Partition the remaining trees into “segments”*). We continue to partition the remainders of each of the K trees by cutting them at each of the joins of branches to form chains from the tree. We will easily be able to track the number of chains produced by this procedure. A second stopping time argument is then applied to cut the chains into “segments” in order to control the number of significant elements added to the approximant (at most $N_0 + 1$ from each segment) and to guarantee that the cumulative effect of the left-over elements (i.e., error) can be controlled by the final Step 5.
- Step 5.** (*Rewrite the “segments” to control error*). Here each segment is rewritten at its finest level, and its terminal element (with the new coefficients) and some of its neighboring elements are added to the approximant. This allows for a void to be created, so that the residual of the segment will have disjoint support with all remaining segments as well as the residuals of those previously processed. This insures that the cumulative pointwise error remains under control.

We now describe these rather vague steps in more detail. Step 1 is clear from our earlier discussion.

Step 2. In the following lemma, we construct a procedure for coloring the elements of Θ with K colors ν , so that no two Courant elements of the same color from the

same level have supports that intersect; in fact, corresponding cells of the same color will have a tree structure with set inclusion as the order relation. This allows us to partition Θ into a disjoint union of sets Θ^ν ($1 \leq \nu \leq K$), and correspondingly organize f as the sum $f = \sum_{\nu=1}^K f_\nu$, where $f_\nu := \sum_{\theta \in \Theta^\nu} b_\theta \varphi_\theta$. We can then proceed to process each of the f_ν without worrying about its terms from the same level overlapping, and at worst a factor of K will come into the constants for the estimates that we derive. For its proof, see the Appendix.

Lemma 3.4 (Coloring lemma). *For any multilevel-triangulation \mathcal{T} of E , the set $\Theta := \Theta(\mathcal{T})$ of all cells generated by \mathcal{T} can be represented as a finite disjoint union of its subsets $(\Theta^\nu)_{\nu=1}^K$ with $K = K(N_0, M_0)$ (N_0 is the maximal valence and M_0 is the maximal number of children of a triangle in \mathcal{T}) such that each Θ^ν has a tree structure with respect to set inclusion, i.e., if $\theta', \theta'' \in \Theta^\nu$ with $(\theta')^\circ \cap (\theta'')^\circ \neq \emptyset$, then either $\theta' \subset \theta''$ or $\theta'' \subset \theta'$.*

In order to complete the remaining Steps 3-5 we must consider two variations in the details of the algorithm, depending on whether $p = \infty$ or $0 < p < \infty$. The case of the uniform metric is presented in Subsection 3.2.1, while the case of L_p ($0 < p < \infty$) is given in Subsection 3.2.2.

3.2.1. *The case $p = \infty$.* Fix $\varepsilon > 0$ and let $\varepsilon^* := \frac{\varepsilon}{2K}$, where we recall that K is the number of colors representing the tree structures.

Step 3. *Trimming of Θ^ν ($1 \leq \nu \leq K$) with ε^* .* We trim each Θ^ν , starting from the finest level Θ_f^ν and proceeding to the coarsest level. We remove from Θ^ν every cell θ° such that

$$(3.5) \quad \sum_{\theta \subset \theta^\circ} |b_\theta| \leq \varepsilon^*.$$

We denote by Γ^ν the set of all $\theta \in \Theta^\nu$ that have been retained after completing this procedure, and by Γ_f^ν the set of all final cells in Γ^ν , i.e., $\theta^\circ \in \Gamma_f^\nu$ iff there is no $\theta \in \Gamma^\nu$ such that $\theta \subsetneq \theta^\circ$. Clearly, for each $\theta^\circ \in \Gamma_f^\nu$,

$$(3.6) \quad \sum_{\theta \subset \theta'} |b_\theta| \leq \varepsilon^* \text{ for each } \theta' \subsetneq \theta^\circ, \text{ but } \sum_{\theta \subset \theta^\circ} |b_\theta| > \varepsilon^*.$$

We denote $f_{\Gamma^\nu} := \sum_{\theta \in \Gamma^\nu} b_\theta \varphi_\theta$. Therefore,

$$(3.7) \quad \|f_\nu - f_{\Gamma^\nu}\|_\infty \leq \max_{\theta^\circ \notin \Gamma^\nu} \left\| \sum_{\theta \subset \theta^\circ} b_\theta \varphi_\theta \right\|_\infty \leq \max_{\theta^\circ \notin \Gamma^\nu} \sum_{\theta \subset \theta^\circ} |b_\theta| \leq \varepsilon^*,$$

and hence, if we set $f_\Gamma := \sum_{\nu=1}^K f_{\Gamma^\nu}$, then

$$(3.8) \quad \|f - f_\Gamma\|_\infty \leq K\varepsilon^* = \varepsilon/2.$$

Step 4. *Partitioning the branches of each tree Γ^ν into chains and the chains into "segments".* For each of the tree structures Γ^ν ($1 \leq \nu \leq K$), we denote by Γ_b^ν the set of all *branching* cells in Γ^ν (cells with more than one child in Γ^ν) and by Γ_{ch}^ν the set of all *chain* cells in Γ^ν (cells with exactly one child in Γ^ν). It is easy to see that

$$(3.9) \quad \#\Gamma_b^\nu \leq \#\Gamma_f^\nu.$$

In fact, one proceeds by induction from the finest to coarser levels, associating each branch cell from Γ_b^ν by a cell from Γ_f^ν . For each branch cell, there is always at least

one member of Γ_f^ν still available from each descendant edge. Only one is used to associate with the current branch cell, thereby leaving at least one available for its next ancestor branch cell in that line.

On the other hand, $\#\Gamma_{ch}^\nu$ may be much larger than $\#\Gamma_f^\nu$, and so we will need to process these elements. A collection of cells $\theta_1 \supset \theta_2 \supset \cdots \supset \theta_l$ is called a *chain* if for $j = 1, \dots, l - 1$, θ_{j+1} is a child of θ_j and $\theta_j \in \Gamma_{ch}^\nu$, and the terminal cell $\theta_l \in \Gamma_f^\nu \cup \Gamma_b^\nu$. We partition the tree Γ^ν into chains. Namely, we start at the coarsest level and construct (maximal) chains which will terminate with either a final cell (in Γ_f^ν) or a branching cell (in Γ_b^ν). We continue this procedure to the finest level.

We next “section” each chain into *segments* using ε^* as a threshold. Namely, if λ is a chain and $\lambda = (\theta_j)_{j=1}^l$ with $\theta_1 \supset \theta_2 \supset \cdots \supset \theta_l$, then we start from the coarsest element θ_1 and sum the coefficients of each cell, moving to the next child of the chain until the sum exceeds the threshold. At this point we cut the chain to form the first (significant) segment and start this procedure again with the next child in line until this is not possible (i.e., ending without the threshold being crossed). We call this type of segment a “remnant segment”. Therefore, this procedure cuts λ into disjoint segments σ of the form $(\theta_j)_{j=i}^{i+\mu}$, $\mu \geq 0$, so that each segment satisfies exactly one of the following conditions:

(a) σ consists of a single “significant cell”:

$$(3.10) \qquad |b_{\theta_i}| > \varepsilon^* \text{ (case of } \mu = 0),$$

(b) σ is a “significant segment”:

$$(3.11) \qquad \sum_{j=i}^{i+\mu-1} |b_{\theta_j}| \leq \varepsilon^*, \text{ but } \sum_{j=i}^{i+\mu} |b_{\theta_j}| > \varepsilon^* \text{ (case of } \mu > 0),$$

(c) σ is a “remnant segment”:

$$(3.12) \qquad \sum_{j=i}^l |b_{\theta_j}| \leq \varepsilon^*.$$

We denote by Σ^ν the set of all such segments $\sigma = (\theta_j)_{j=i}^{i+\mu}$ resulting from this procedure.

Step 5. *Rewriting elements from certain segments of Σ^ν .* Let $\sigma = (\theta_j)_{j=1}^\mu$ be any segment from Σ^ν , and suppose that the finest cell θ_μ of σ belongs to Θ_m . We rewrite the Courant elements $(\sum_{j=1}^\mu b_{\theta_j} \varphi_{\theta_j})$ of the segment at its finest (m -th) level, finding coefficients (c_θ) such that

$$\sum_{\theta \in \Theta_m, \theta^\circ \cap \theta_\mu \neq \emptyset} c_\theta \varphi_\theta = \sum_{j=1}^\mu b_{\theta_j} \varphi_{\theta_j} \text{ on } \theta_\mu.$$

We denote $\mathcal{X}_\sigma := \{\theta \in \Theta_m : \theta^\circ \cap \theta_\mu \neq \emptyset \text{ and } \theta \subset \theta_1\}$. Obviously, if $\mu = 1$ (i.e., the segment consists of a single cell), then the coefficient remains unchanged and $\mathcal{X}_\sigma = \sigma = \{\theta_1\}$. Observe in any case that $\#\mathcal{X}_\sigma \leq N_0 + 1$ and $\bigcup_{\theta \in \mathcal{X}_\sigma} \theta \subset \theta_1$. Finally, set $\Sigma := \bigcup_{\nu=1}^K \Sigma^\nu$, and correspondingly define

$$(3.13) \qquad A_\varepsilon^{TC}(f) := \sum_{\sigma \in \Sigma} \sum_{\theta \in \mathcal{X}_\sigma} c_\theta \varphi_\theta$$

as our approximant produced by the “trim and cut” algorithm.

• **Error estimation for the “trim and cut” algorithm (case $p = \infty$).** Suppose that the “trim and cut” procedure has been applied to a function f with $\varepsilon > 0$, and $A_\varepsilon^{TC}(f) = \sum_{\theta \in \Lambda_\varepsilon} c_\theta \varphi_\theta$ is the resulting approximant from (3.13), where $\Lambda_\varepsilon = \bigcup_{\sigma \in \Sigma} \mathcal{X}_\sigma$. We denote

$$n(\varepsilon) := n_f(\varepsilon) := \#\Lambda_\varepsilon, \quad \mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty := \|f - A_\varepsilon^{TC}(f)\|_\infty,$$

and

$$\mathbb{A}_n^{TC}(f)_\infty := \inf \left\{ \mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty : n(\varepsilon) \leq n \right\}.$$

Note that each of these quantities depend implicitly on \mathcal{T} . To complete our results for the “trim and cut” algorithm, we show first in Lemma 3.5 that this is a good approximation to f , and then that the number of elements that are used in the approximant satisfies the correct estimates (see Theorem 3.7 below).

Lemma 3.5. *Suppose that $A_\varepsilon^{TC}(f)$ is the approximant for f given in equation (3.13) which has been constructed using the “trim and cut” algorithm. Then*

$$(3.14) \quad \|f - A_\varepsilon^{TC}(f)\|_\infty \leq \varepsilon.$$

Proof. Following the definition (3.13) of $A_\varepsilon^{TC}(f)$, we define

$$A^\nu := \sum_{\sigma' \in \Sigma^\nu} \sum_{\theta \in \mathcal{X}_{\sigma'}} c_\theta \varphi_\theta.$$

Then obviously $A_\varepsilon^{TC}(f) = \sum_{\nu=1}^K A^\nu$. Since $\varepsilon^* = \frac{\varepsilon}{2K}$, it suffices to show that $\|f_{\Gamma^\nu} - A^\nu\|_\infty \leq \varepsilon^*$.

In Step 5 we extracted the heart of each segment $\sigma = (\theta_j)_{j=1}^\mu$, added its contribution to the approximant (3.13), and cleared room for descendant cells. To estimate the associated error, we introduce the *ring* for σ as $R_\sigma := \theta_1 \setminus \theta_\mu$; then $R_\sigma = \emptyset$ when σ consists of a significant cell (i.e., condition (3.10) holds). For any nonempty ring R_σ ($\sigma \in \Sigma^\nu$), set $\sigma' := (\theta_j)_{j=1}^{\mu-1}$ and observe that at worst

$$(3.15) \quad \begin{aligned} \|f_{\Gamma^\nu} - A^\nu\|_{L_\infty(R_\sigma)} &= \left\| \sum_{\theta \in \sigma} b_\theta \varphi_\theta - \sum_{\theta \in \mathcal{X}_{\sigma'}} c_\theta \varphi_\theta \right\|_{L_\infty(R_\sigma)} \\ &\leq \left\| \sum_{\theta \in \sigma'} b_\theta \varphi_\theta \right\|_{L_\infty(\theta_1)} \leq \sum_{\theta \in \sigma'} |b_\theta| \leq \varepsilon^*. \end{aligned}$$

It is easy to see that all rings R_σ ($\sigma \in \Sigma^\nu$) are disjoint and the set where A^ν may differ from f_{Γ^ν} is contained in $\bigcup_{\sigma \in \Sigma^\nu} R_\sigma$. Hence, by summing over all segments σ and then over all colors ν , it follows that

$$\|f_\Gamma - A_\varepsilon^{TC}(f)\|_\infty \leq \sum_{\nu=1}^K \sum_{\sigma \in \Sigma^\nu} \|f_{\Gamma^\nu} - A^\nu\|_{L_\infty(R_\sigma)} \leq K\varepsilon^* = \frac{\varepsilon}{2}.$$

This together with estimate (3.8) implies the desired error estimate (3.14). □

Remark 3.6. Conditions (3.5), (3.11), and (3.12) can be relaxed by replacing every sum $\sum |b_\theta|$ by $\|\sum b_\theta \varphi_\theta\|_\infty$. This would not change the rate of approximation, but may improve the constants in a practical implementation.

Theorem 3.7. *If $f \in B_\tau^\alpha(\mathcal{T})$, $\alpha \geq 1$, $\tau := 1/\alpha$, then for each $\varepsilon > 0$,*

$$(3.16) \quad \mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty \leq \varepsilon \text{ and } n(\varepsilon) \leq c \varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau,$$

where $c = c(N_0, M_0, \alpha)$. Therefore,

$$(3.17) \quad \mathbb{A}_n^{TC}(f)_\infty \leq c n^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})}.$$

Proof. We have already shown in Lemma 3.5 that $\mathbb{A}_{n(\varepsilon)}^{TC}(f)_\infty \leq \varepsilon$; so we only need to establish $n(\varepsilon) \leq c \varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau$. We first observe that it is enough to estimate $\#\Sigma^\nu$, since contributions to the approximant occur only as each segment from Σ^ν is processed. Note that at most one element is contributed for segments consisting of a single significant cell (3.10) and at most $N_0 + 1$ contributions for the segments satisfying instead either (3.11) or (3.12).

In order to estimate $\#\Sigma^\nu$ we first estimate $\#\Gamma_f^\nu$, since it will estimate certain terms. The stopping criterium (3.6) in Step 3,

$$(3.18) \quad \varepsilon^* < \sum_{\theta \subseteq \theta^\diamond} |b_\theta|,$$

must hold for each $\theta^\diamond \in \Gamma_f^\nu$. So if we apply the τ -th power to both sides, use the embedding of the sequence spaces ($\tau \leq 1$), sum over all $\theta \in \Gamma_f^\nu$, and observe that the supports of the cells in Γ_f^ν have disjoint interiors, then we obtain

$$(3.19) \quad \#\Gamma_f^\nu (\varepsilon^*)^\tau < \sum_{\theta^\diamond \in \Gamma_f^\nu} \sum_{\theta \subseteq \theta^\diamond} |b_\theta|^\tau \leq \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau.$$

The rightmost inequality follows immediately by our definition of the norm of $B_\tau^\alpha(\mathcal{T})$ (see (3.2)).

To complete the proof of the theorem, we only need to establish a similar estimate for the number of elements of Σ^ν . Recall, however, that the segments σ are formed as disjoint segments of cells from the tree structure and come as one of two types, Σ_{sig} , those exceeding the threshold (see conditions (3.10) or (3.11)) and, Σ_{rem} , those that do not (see condition (3.12)). From the construction it follows that remnant segments terminate with either a unique final cell or a unique branching cell, and so by (3.9),

$$(3.20) \quad \#\Sigma_{\text{rem}} \leq \#\Gamma_b^\nu + \#\Gamma_f^\nu \leq 2 \#\Gamma_f^\nu,$$

which has just been shown in (3.19) to satisfy the desired bound.

Therefore we are reduced to estimating $\#\Sigma_{\text{sig}}$. But the same idea used in estimating $\#\Gamma_f^\nu$ (see (3.18)–(3.19)) may be employed once again. Indeed, we just replace the condition (3.18) with

$$(3.21) \quad \varepsilon^* < \sum_{\theta \in \sigma} |b_\theta|,$$

and use the fact that the segments are disjoint (considered as part of the tree structure), in order to obtain

$$(3.22) \quad \#\Sigma_{\text{sig}} (\varepsilon^*)^\tau < \sum_{\sigma \in \Sigma_{\text{sig}}} \sum_{\theta \in \sigma} |b_\theta|^\tau \leq \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau. \quad \square$$

Although not required here, the following lemma will be needed in §5 and can now be established using the techniques of this section.

Lemma 3.8. *Let $f = f^0 + f^1$, where $f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta$, $f^j = \sum_{\theta \in \Theta} b_\theta^j \varphi_\theta$ ($j = 0, 1$) with $b_\theta = b_\theta^0 + b_\theta^1$ (all $\theta \in \Theta$), and let*

$$\mathcal{N}_j := \left(\sum_{\theta \in \Theta} |b_\theta^j|^{\tau_j} \right)^{1/\tau_j} < \infty \quad (j = 0, 1)$$

with $\alpha_j \geq 1$ and $\tau_j = 1/\alpha_j$. If the “trim and cut” algorithm with $\varepsilon = \varepsilon_0 + \varepsilon_1$ ($\varepsilon_j > 0$) has been applied to f , represented as above in place of Step 1, then

$$(3.23) \quad \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^{TC}(f)_\infty \leq \varepsilon_0 + \varepsilon_1,$$

$$(3.24) \quad n(\varepsilon_0 + \varepsilon_1) \leq c\varepsilon^{-\tau_0} \mathcal{N}_0^{\tau_0} + c\varepsilon^{-\tau_1} \mathcal{N}_1^{\tau_1},$$

and consequently

$$(3.25) \quad \mathbb{A}_n^{TC}(f)_\infty \leq cn^{-\alpha_0} \mathcal{N}_0 + cn^{-\alpha_1} \mathcal{N}_1, \quad n = 1, 2, \dots,$$

with c depending only on α_0, α_1 , and the parameters of \mathcal{T} .

Proof. All the elements for the proof already appear in this subsection, especially in the proofs of Theorem 3.7 and Lemma 3.5, and we shall assume complete familiarity with the notation, terminology, and estimates given there. Denote the number of cells used in the “trim and cut” algorithm for (b_θ) , with approximation error ε , by $n(\varepsilon)$. Similarly, let $n_j(\varepsilon_j)$ be the corresponding number of cells used for f^j ($j = 0, 1$), again represented as $f^j = \sum_{\theta \in \Theta} b_\theta^j \varphi_j$, in place of Step 1. The theorem will be proved once we establish the estimate

$$(3.26) \quad n(\varepsilon_0 + \varepsilon_1) \leq 2(n_0(\varepsilon_0) + n_1(\varepsilon_1))$$

for any $\varepsilon_0, \varepsilon_1 > 0$. Indeed, by combining this inequality with the results of Theorem 3.7 (in particular, inequalities (3.16)–(3.17)), we can see that the estimate

$$(3.27) \quad n(\varepsilon_0 + \varepsilon_1) \leq 2c\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + 2c\varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} = n$$

is true if we set $\varepsilon_j := (4c)^{1/\tau_j} n^{-1/\tau_j} \mathcal{N}_j$, $j = 0, 1$, where c is the constant appearing there. But the fact that $n \geq n(\varepsilon_0 + \varepsilon_1)$ and the definition of $n(\cdot)$ imply

$$\mathbb{A}_n^{TC}(f)_\infty \leq \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^{TC}(f)_\infty \leq \varepsilon_0 + \varepsilon_1.$$

Hence, by the definition of the ε_j , the rightmost terms of this last inequality are bounded by the desired terms on the right-hand side of inequality (3.25).

In order to prove estimate (3.26), we only need to estimate the number of segments Σ for f . First observe in Step 3 of the algorithm that for the thresholding condition (3.6) to hold for f , with $\varepsilon := \varepsilon_0 + \varepsilon_1$, the condition must also be satisfied for that same cell θ^\diamond for at least one of the f^j with corresponding threshold ε_j ($j = 0, 1$). This shows that the tree $\Gamma^\nu = \Gamma^\nu(f, \varepsilon)$ determined by threshold ε is contained in the union of the corresponding trees $\Gamma^\nu(f^j, \varepsilon_j)$ ($j = 0, 1$). By the construction of segments σ from maximal chains of $\Gamma^\nu(f)$ in Step 4, the segments for f are disjoint and one of the conditions (3.10)–(3.12) must hold. If (3.10) or (3.11) holds for a segment σ of f , then $\sum_{\theta \in \sigma} |b_\theta^0 + b_\theta^1| > \varepsilon_0 + \varepsilon_1$ implies the corresponding condition for at least one of f^0 (and ε_0) or f^1 (and ε_1). That is, for one of $j = 0, 1$ we must have $\sum_{\theta \in \sigma} |b_\theta^j| > \varepsilon_j$, and so for at least half of the segments of f this condition must persist for a fixed index j ($j = 0, 1$). The number of remnant segments (see (3.12)), on the other hand, may be estimated by the sum of the number of remnant segments of f^0 and f^1 , plus the number of *new* branching cells which may arise within the union of the trees of f^0 and f^1 . These new cells are

introduced in $\Gamma^\nu(f, \varepsilon)$ when two chains, exclusive to each of the $\Gamma^\nu(f^j, \varepsilon_j)$, meet, thereby dividing the existing chains for each of the trees and creating an additional segment. It is easy to see that the number of such new branching cells does not exceed $\min\{\Gamma_f^\nu(f^0, \varepsilon_0), \Gamma_f^\nu(f^1, \varepsilon_1)\}$.

This accounting of the three qualifying conditions (3.10)-(3.12) for segments gives

$$\begin{aligned} \left\lceil \frac{\#\Sigma}{2} \right\rceil &\leq \max\{\#\Sigma(f^0, \varepsilon_0), \#\Sigma(f^1, \varepsilon_1)\} + \min\{\Gamma_f^\nu(f^0, \varepsilon_0), \Gamma_f^\nu(f^1, \varepsilon_1)\} \\ &\leq \#\Sigma(f^0, \varepsilon_0) + \#\Sigma(f^1, \varepsilon_1), \end{aligned}$$

which implies the desired estimate (3.26) and completes the proof. \square

3.2.2. *The case $0 < p < \infty$.* We now return to completing Steps 3-5 in the case that $p < \infty$. The arguments are quite similar to the case $p = \infty$ in the previous subsection, and we shall use the notation there and indicate only the differences. Introduce a new parameter ϱ , where $0 < \varrho < p$, and fix $\varepsilon > 0$.

Step 3. *Trimming of Θ^ν ($1 \leq \nu \leq K$) with ε .* This step is the same as in Case 1 ($p = \infty$) with (3.5) replaced by

$$(3.28) \qquad \qquad \qquad \left(\sum_{\theta \in \theta^\circ} (|b_\theta| |\theta|^{1/p})^\varrho \right)^{1/\varrho} \leq \varepsilon.$$

In contrast to the case $p = \infty$, the error $\|f_\nu - f_{\Gamma^\nu}\|_p$ is no longer controlled solely by ε . It will depend on the smoothness of the function f that is being approximated (see Theorem 3.9 below).

Step 4. *Partitioning the branches of each tree Γ^ν into chains and the chains into “segments”.* We proceed exactly as in the case $p = \infty$, replacing conditions (3.10)-(3.12) by the following:

$$(3.29) \qquad \qquad \qquad |b_{\theta_i}| |\theta_i|^{1/p} > \varepsilon \quad (\text{case of } \mu = 0),$$

$$(3.30) \qquad \qquad \qquad \left(\sum_{j=i}^{i+\mu-1} (|b_{\theta_j}| |\theta_j|^{1/p})^\varrho \right)^{1/\varrho} \leq \varepsilon, \text{ but } \left(\sum_{j=i}^{i+\mu} (|b_{\theta_j}| |\theta_j|^{1/p})^\varrho \right)^{1/\varrho} > \varepsilon \quad (\text{case of } \mu > 0),$$

$$(3.31) \qquad \qquad \qquad \left(\sum_{j=i}^l (|b_{\theta_j}| |\theta_j|^{1/p})^\varrho \right)^{1/\varrho} \leq \varepsilon.$$

Step 5. *Rewriting elements from certain segments of Σ^ν .* This step is exactly the same as for the case $p = \infty$.

• **Error estimation for the “trim and cut” algorithm (case $0 < p < \infty$).**

Suppose that the “trim and cut” algorithm has been applied to a function f with $0 < \varrho < p$ and $\varepsilon > 0$, as described above. Let $A_\varepsilon^{TC}(f)_p = \sum_{\theta \in \Lambda_\varepsilon} c_\theta \varphi_\theta$, $\Lambda_\varepsilon \subset \Theta$, be the approximant produced by the algorithm. We denote

$$n(\varepsilon) := \#\Lambda_\varepsilon, \qquad \mathbb{A}_{n(\varepsilon)}^{TC}(f)_p := \|f - A_\varepsilon^{TC}(f)_p\|_p,$$

and

$$\mathbb{A}_n^{TC}(f)_p := \inf\{\mathbb{A}_{n(\varepsilon)}^{TC}(f)_p : n(\varepsilon) \leq n\}.$$

Theorem 3.9. *If $f \in B_\tau^\alpha(\mathcal{T})$, where $\alpha \geq 1/\varrho - 1/p$ and $\tau = (\alpha + \frac{1}{p})^{-1}$, then for each $\varepsilon > 0$,*

$$(3.32) \quad \mathbb{A}_{n(\varepsilon)}^{TC}(f)_p \leq c\varepsilon^{\alpha\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^{\tau/p} \quad \text{and} \quad n(\varepsilon) \leq c\varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau,$$

and hence

$$(3.33) \quad \mathbb{A}_n^{TC}(f)_p \leq cn^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})}, \quad n = 1, 2, \dots,$$

where c depends on p, ϱ, α , and the parameters of \mathcal{T} .

Proof. We first estimate $n(\varepsilon)$. From the stopping time criterium (the converse inequality of (3.28)) in Step 3, it follows that

$$(3.34) \quad \varepsilon < \left(\sum_{\theta \subset \theta^\diamond} (|b_\theta| |\theta|^{1/p})^\varrho \right)^{1/\varrho} \leq \left(\sum_{\theta \subset \theta^\diamond} (|b_\theta| |\theta|^{1/p})^\tau \right)^{1/\tau} \quad (\text{since } \tau \leq \varrho)$$

for each $\theta^\diamond \in \Gamma_f^\nu$, which enables us to repeat the arguments from the proof of Theorem 3.7 and obtain the estimate $\#\Gamma_f^\nu \leq c\varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau$. In going further, we use (3.30) in a similar fashion and the above to infer as in the proof of Theorem 3.7 that

$$(3.35) \quad \#\Sigma^\nu \leq c\varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau.$$

This implies the desired estimate for $n(\varepsilon)$.

It remains to estimate the error $\|f - A_{n(\varepsilon)}^{TC}(f)_p\|_p$. We first estimate $\|f_\nu - f_{\Gamma^\nu}\|_p$. To this end, we group the removed cells into collections of comparable B_τ^α -norms. We denote by

$$\Xi^\nu := \{\theta \in \Theta^\nu \setminus \Gamma^\nu : \theta \not\subseteq \theta' \text{ for any } \theta' \in \Theta^\nu \setminus \Gamma^\nu, \theta' \neq \theta\}$$

the set of all cells at which a trimmed branch starts. Note that for each $\theta^\diamond \in \Xi^\nu$ the inequality (3.28) holds. Therefore, we can partition Ξ^ν into disjoint collections Ξ_j^ν , $j = 1, 2, \dots, L^\nu$, such that $\Xi^\nu = \bigcup_{j=1}^{L^\nu} \Xi_j^\nu$ and

$$(3.36) \quad \varepsilon^\varrho < \sum_{\theta^\diamond \in \Xi_j^\nu} \sum_{\theta \subset \theta^\diamond} (|b_\theta| |\theta|^{1/p})^\varrho \leq 2\varepsilon^\varrho$$

for all $j = 1, 2, \dots, L^\nu$ except possibly for $j = L^\nu$, when the leftmost inequality may fail. Hence, since the cells from Ξ^ν have disjoint interiors, and recalling that $|b_\theta| |\theta|^{1/p} \approx \|b_\theta \varphi_\theta\|_p$, we obtain

$$\begin{aligned} \|f_\nu - f_{\Gamma^\nu}\|_p &= \left\| \sum_{j=1}^{L^\nu} \sum_{\theta^\diamond \in \Xi_j^\nu} \sum_{\theta \subset \theta^\diamond} b_\theta \varphi_\theta \right\|_p \leq \left(\sum_{j=1}^{L^\nu} \left\| \sum_{\theta^\diamond \in \Xi_j^\nu} \sum_{\theta \subset \theta^\diamond} b_\theta \varphi_\theta \right\|_p^p \right)^{\frac{1}{p}} \\ (3.37) \quad &\leq c \left(\sum_{j=1}^{L^\nu} \left[\sum_{\theta^\diamond \in \Xi_j^\nu} \sum_{\theta \subset \theta^\diamond} (|b_\theta| |\theta|^{1/p})^\varrho \right]^{p/\varrho} \right)^{\frac{1}{p}} \\ &\leq c \left(\sum_{j=1}^{L^\nu} 2^{p/\varrho} \varepsilon^p \right)^{1/p} = c(L^\nu)^{1/p} \varepsilon, \end{aligned}$$

where we used the embedding inequality (2.26). To estimate L^ν we once again exploit the idea used in estimating $\#\Gamma_f^\nu$ (see (3.18)–(3.19)). Since $0 < \tau \leq \varrho$, we

have by (3.36) that

$$\varepsilon < \left(\sum_{\theta^\circ \in \Xi_j^\nu} \sum_{\theta \subset \theta^\circ} (|b_\theta| |\theta|^{1/p})^q \right)^{1/q} \leq \left(\sum_{\theta^\circ \in \Xi_j^\nu} \sum_{\theta \subset \theta^\circ} (|b_\theta| |\theta|^{1/p})^\tau \right)^{1/\tau}.$$

We use this and the fact that the collections Ξ_j^ν are disjoint to obtain

$$(3.38) \quad L^\nu \cdot \varepsilon^\tau \leq c \sum_{j=1}^{L^\nu} \left(\sum_{\theta^\circ \in \Xi_j^\nu} \sum_{\theta \subset \theta^\circ} \|b_\theta \varphi_\theta\|_p^\tau \right) \leq c \|f_\nu\|_{B_\tau^\alpha(T)}^\tau.$$

Combining (3.37) and (3.38), we obtain

$$\|f_\nu - f_{\Gamma^\nu}\|_p \leq c (\varepsilon^{-\tau} \|f_\nu\|_{B_\tau^\alpha(T)}^\tau)^{1/p} \varepsilon = c \varepsilon^{\alpha\tau} \|f_\nu\|_{B_\tau^\alpha(T)}^{\tau/p},$$

and hence by standard subadditivity estimates for L_p ($0 < p < \infty$) we may estimate the sum

$$(3.39) \quad \begin{aligned} \|f - f_\Gamma\|_p &\leq \left(\sum_{\nu=1}^K \|f_\nu - f_{\Gamma^\nu}\|_p^{p^*} \right)^{1/p^*} \\ &\leq c \varepsilon^{\alpha\tau} \left(\sum_{\nu=1}^K (\|f_\nu\|_{B_\tau^\alpha(T)}^\tau)^{p^*/p} \right)^{1/p^*} \leq c \varepsilon^{\alpha\tau} \|f\|_{B_\tau^\alpha(T)}^{\tau/p}, \end{aligned}$$

where $p^* := \min\{1, p\}$.

To complete the proof of the theorem, we must estimate $\|f_{\Gamma^\nu} - A^\nu\|_p$. This differs from our earlier arguments in the case $p = \infty$, which involved the error estimate (3.15) over a ring of a segment. For any such ring R_σ ($\sigma \in \Sigma^\nu$) we use instead the estimate

$$\begin{aligned} \|f_{\Gamma^\nu} - A^\nu\|_{L_p(R_\sigma)} &= \left\| \sum_{\theta \in \sigma} b_\theta \varphi_\theta - \sum_{\theta \in \mathcal{X}_\sigma} c_\theta \varphi_\theta \right\|_{L_p(R_\sigma)} \leq \left\| \sum_{\theta \in \sigma'} b_\theta \varphi_\theta \right\|_{L_p(\theta_1)} \\ &\leq c \left(\sum_{\theta \in \sigma'} \|b_\theta \varphi_\theta\|_p^\tau \right)^{1/\tau} \leq c \left(\sum_{\theta \in \sigma'} (|b_\theta| |\theta|^{1/p})^q \right)^{1/q} \leq c \varepsilon, \end{aligned}$$

where we used the embedding inequality (2.26). From the above, using that all rings $\{R_\sigma\}_{\sigma \in \Sigma^\nu}$ have disjoint interiors, we obtain

$$(3.40) \quad \|f_{\Gamma^\nu} - A^\nu\|_p \leq \left(\sum_{\sigma \in \Sigma^\nu} \|f_{\Gamma^\nu} - A^\nu\|_{L_p(R_\sigma)}^p \right)^{1/p} \leq c (\#\Sigma^\nu)^{1/p} \varepsilon.$$

Combining (3.40) and (3.35) yields

$$\|f_{\Gamma^\nu} - A^\nu\|_p \leq c \varepsilon^{\alpha\tau} \|f_\nu\|_{B_\tau^\alpha(T)}^{\tau/p},$$

and hence

$$\begin{aligned} \|f_\Gamma - A_\varepsilon^{TC}(f)_p\|_p &\leq \left(\sum_{\nu=1}^K \|f_{\Gamma^\nu} - A^\nu\|_p^{p^*} \right)^{1/p^*} \\ &\leq c \varepsilon^{\alpha\tau} \left(\sum_{\nu=1}^K (\|f_\nu\|_{B_\tau^\alpha(T)}^\tau)^{p^*/p} \right)^{1/p^*} \leq c \varepsilon^{\alpha\tau} \|f\|_{B_\tau^\alpha(T)}^{\tau/p}, \end{aligned}$$

where $p^* := \min\{1, p\}$. From this and (3.39), we obtain the appropriate estimate which corresponds to (3.14) of the case for $p = \infty$:

$$(3.41) \quad \|f - A_{\varepsilon}^{TC}(f)_p\|_p \leq c \varepsilon^{\alpha\tau} \|f\|_{B_{\tau}^{\alpha}(\mathcal{T})}^{\tau/p}. \quad \square$$

Lemma 3.10. *Let $f = f^0 + f^1$, where $f = \sum_{\theta \in \Theta} b_{\theta} \varphi_{\theta}$, $f^j = \sum_{\theta \in \Theta} b_{\theta}^j \varphi_{\theta}$ ($j = 0, 1$) with $b_{\theta} = b_{\theta}^1 + b_{\theta}^2$ (all θ), and let*

$$\mathcal{N}_j := \left(\sum_{\theta \in \Theta} (|b_{\theta}^j| |\theta|^{1/p})^{\tau_j} \right)^{1/\tau_j} < \infty \quad (j = 0, 1)$$

with $\alpha_j \geq \frac{1}{\varrho} - \frac{1}{p}$ ($0 < \varrho < p$) and $\tau_j := 1/(\alpha_j + \frac{1}{p})^{-1}$, $j = 0, 1$. Furthermore, suppose the “trim and cut” algorithm has been applied to f , using the above representation of f in place of Step 1, with $0 < \varrho < p$ as above and $\varepsilon = \varepsilon_0 + \varepsilon_1$ for some $\varepsilon_0, \varepsilon_1 > 0$. Then we have

$$(3.42) \quad \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^{TC}(f)_p \leq c (\varepsilon_0 + \varepsilon_1) \left(\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + \varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} \right)^{1/p},$$

$$(3.43) \quad n(\varepsilon_0 + \varepsilon_1) \leq c \left(\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + \varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} \right),$$

and, therefore,

$$(3.44) \quad \mathbb{A}_n^{TC}(f)_p \leq c \left(n^{-\alpha_0} \mathcal{N}_0 + n^{-\alpha_1} \mathcal{N}_1 \right), \quad n = 1, 2, \dots,$$

where c depends only on $p, \varrho, \alpha_0, \alpha_1$, and the parameters of \mathcal{T} .

Proof. The proof is very similar to the proof of Theorem 3.9, and we shall only indicate the differences, using the notation and ideas from there. Those differences are in estimating $\#\Gamma_f^{\nu}$, $\#\Sigma^{\nu}$ and L^{ν} (see (3.35) and (3.38)). From the stopping criterium (converse inequality to (3.28)) in Step 3, it follows that, for $\theta^{\circ} \in \Gamma_f^{\nu}$,

$$\begin{aligned} \varepsilon_0 + \varepsilon_1 &< \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}| |\theta|^{1/p})^{\varrho} \right)^{1/\varrho} \\ &\leq c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^0| |\theta|^{1/p})^{\varrho} \right)^{1/\varrho} + c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^1| |\theta|^{1/p})^{\varrho} \right)^{1/\varrho} \\ &\leq c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^0| |\theta|^{1/p})^{\tau_0} \right)^{1/\tau_0} + c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^1| |\theta|^{1/p})^{\tau_1} \right)^{1/\tau_1}, \end{aligned}$$

where $c_{\varrho} := \max\{1, 2^{1/\varrho-1}\}$ and we used the fact that $\tau_0, \tau_1 \leq \varrho$. Therefore, for each $\theta^{\circ} \in \Gamma_f^{\nu}$, at least one of

$$\varepsilon_0 < c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^0| |\theta|^{1/p})^{\tau_0} \right)^{1/\tau_0} \quad \text{or} \quad \varepsilon_1 < c_{\varrho} \left(\sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^1| |\theta|^{1/p})^{\tau_1} \right)^{1/\tau_1}$$

must hold. Denoting by $\Gamma_{f^0}^{\nu}$ and $\Gamma_{f^1}^{\nu}$ the sets of all $\theta^{\circ} \in \Gamma_f^{\nu}$ for which the first or second inequality, respectively, holds, we obtain

$$\mathcal{N}_j^{\tau_j} \geq c \sum_{\theta^{\circ} \in \Gamma_{f^j}^{\nu}} \sum_{\theta \subset \theta^{\circ}} (|b_{\theta}^j| |\theta|^{1/p})^{\tau_j} \geq c \#\Gamma_{f^j}^{\nu} \varepsilon_j^{\tau_j} \quad (j = 0, 1),$$

and hence

$$(3.45) \quad \#\Gamma_f^\nu \leq \#\Gamma_{f_0}^\nu + \#\Gamma_{f_1}^\nu \leq c \left(\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + \varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} \right).$$

We obtain similar (with the same right-hand-side quantity) for $\#\Sigma^\nu$ and L^ν by using the same argument. The estimate for $\#\Sigma^\nu$ gives the desired estimate for $n(\varepsilon_0 + \varepsilon_1)$.

We may use estimates (3.37) and (3.40) in the proof of Theorem 3.9, with $\varepsilon = \varepsilon_0 + \varepsilon_1$, together with the above estimates for $\#\Sigma^\nu$ and L^ν , to obtain

$$(3.46) \quad \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^{TC}(f)_p \leq c(\varepsilon_0 + \varepsilon_1) (\#\Sigma)^{1/p},$$

from which the desired estimate (3.42) follows. The final estimate (3.44) is proved by selecting $\varepsilon_j = (2c/n)^{1/\tau_j} \mathcal{N}_j$, which by our result (3.43), gives that $n(\varepsilon_0 + \varepsilon_1) \leq n$ and so

$$\mathbb{A}_n^{TC}(f)_p \leq \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^{TC}(f)_p \leq c n^{1/p} (\varepsilon_0 + \varepsilon_1) \leq c \left(n^{-\alpha_0} \mathcal{N}_0 + n^{-\alpha_1} \mathcal{N}_1 \right),$$

where we have used (3.46) in the second inequality. \square

3.3. “Push the error” algorithm. The idea of this algorithm to our knowledge first appeared in [5]. Our goal is to adapt this algorithm for nonlinear n -term Courant element approximation in the uniform norm and perfect it so that the resulting algorithm achieves the rate of convergence of the best approximation.

In §3.3.1, we describe the “push the error” algorithm in its simplest and most naive form. We follow with three examples which illustrate deficiencies of the simple algorithm and the types of traps to which it may fall prey. In §3.3.2, we give our refined version of that algorithm. Throughout this section, we assume that $\mathcal{T} = \bigcup_{m=0}^\infty \mathcal{T}_m$ is an LR-triangulation of some compact polygonal domain E in \mathbb{R}^2 , where the approximation takes place (see §2.1), and $f \in C(E)$.

3.3.1. A naive “push the error” algorithm ($p = \infty$). We begin by outlining the basic elements of the algorithm.

Step 1 (Decompose). In this subsection we denote by $Q_j(f)$ the piecewise linear continuous function that interpolates f at the vertices V_j of all triangles from \mathcal{T}_j . Clearly $f \in C(E)$ can be represented as follows:

$$(3.47) \quad f = Q_0(f) + \sum_{j=1}^\infty (Q_j(f) - Q_{j-1}(f)) =: \sum_{\theta \in \Theta} c_\theta \varphi_\theta,$$

where the series converges uniformly. In practice the series terminates at some finest level Θ_J ($J > 1$), so that

$$f = \sum_{j=0}^J \sum_{\theta \in \Theta_j} c_\theta \varphi_\theta.$$

Assuming that initially $f = \sum_{\theta \in \Theta_J} c_\theta \varphi_\theta$, there exists a fast and efficient procedure for obtaining (3.47).

Step 2 (“Threshold” and “push the error”). Fix $\varepsilon > 0$. We shall begin at the coarsest level Θ_0 and proceed consecutively through to higher resolution levels

$\Theta_1, \Theta_2, \dots, \Theta_J$. We define Λ_0 as the set of all cells $\theta \in \Theta_0$ such that $|c_\theta| > \varepsilon$ ($\|\varphi_\theta\| = 1$), and set

$$A_0 := \sum_{\theta \in \Lambda_0} c_\theta \varphi_\theta =: \sum_{\theta \in \Theta_0} b_\theta \varphi_\theta.$$

Next we rewrite all remaining terms $c_\theta \varphi_\theta$ ($\theta \in \Theta_0 \setminus \Lambda_0$) at the next finer level and add the resulting terms to the corresponding terms from $(c_\theta \varphi_\theta)_{\theta \in \Theta_1}$. Thus we obtain a representation of f in the form

$$f = A_0 + \sum_{\theta \in \Theta_1} b_\theta \varphi_\theta + \sum_{j=2}^J \sum_{\theta \in \Theta_j} c_\theta \varphi_\theta.$$

We next process the Courant elements at level Θ_1 . We define Λ_1 as the set of all $\theta \in \Theta_1$ such that $|b_\theta| > \varepsilon$, and set $A_1 := \sum_{\theta \in \Lambda_1} b_\theta \varphi_\theta$. All remaining terms $b_\theta \varphi_\theta$, $\theta \in \Theta_1 \setminus \Lambda_1$, we rewrite at the finer level Θ_2 and add the resulting terms to the corresponding terms $(c_\theta \varphi_\theta)_{\theta \in \Theta_2}$. The representation of f at this stage is written as

$$(3.48) \quad f = A_0 + A_1 + \sum_{\theta \in \Theta_2} b_\theta \varphi_\theta + \sum_{j=3}^J \sum_{\theta \in \Theta_j} c_\theta \varphi_\theta.$$

We continue in this way until we reach the finest (i.e., highest resolution) level Θ_J . The only modification at this finest level is that we discard all terms whose coefficients in absolute value do not exceed our threshold parameter ε . In this way we obtain our approximation

$$A := A_\varepsilon(f) := \sum_{j=0}^J A_j = \sum_{\theta \in \Lambda} b_\theta \varphi_\theta, \quad \Lambda := \bigcup_{j=0}^J \Lambda_j.$$

Since only small terms ($|b_\theta| \leq \varepsilon$) at a single (in this case, finest) level are discarded, they cannot stack up, and we have

$$\|f - A_\varepsilon(f)\|_\infty \leq \varepsilon.$$

Some modifications must be made, however, to insure that this simple and efficient algorithm will achieve sparse representations in an asymptotically optimal sense and avoid hidden traps that will result in using too many terms in the approximation.

We indicate briefly each of the possible pitfalls to keep in mind, before developing the algorithm in full in the next subsection.

Trap 1. The interpolation scheme we used to represent f in (3.47) leads to difficulties, since it does not always lead to sparse representations. We give here a univariate example which may be easily extended to two dimensions.

Let $E := [-1, 1]$, and let f be the hat function on $[-\frac{1}{2^N}, \frac{1}{2^N}]$ for N sufficiently large, i.e., $f(x) = \varphi(2^N x)$ with $\varphi(x) := (1 - |x|) \mathbb{1}_{[-1, 1]}(x)$, $x \in \mathbb{R}$. We assume that \mathcal{T} consists of all dyadic subintervals of $[-1, 1]$. Using the interpolation scheme described in Step 1 at the coarsest level, we must interpolate the extremes at $-1, 0, 1$ in order to decrease the L^∞ error. The resulting error after this stage, however, is $1 - \frac{1}{2^N}$. Proceeding with the *naive* “push-the-error” algorithm with any $\varepsilon < \frac{1}{2}$ results in an index set Λ with $\#\Lambda \sim N$. However, the best approximation is achieved using the single fine scale element $\varphi(2^N x)$. Therefore, any reasonable algorithm that retains n terms in the approximant should give a rate of convergence $\mathcal{O}(n^{-\gamma})$ for any $\gamma > 0$.

Trap 2. For a given $\varepsilon > 0$ the algorithm as currently described may produce a great number of undesired terms due to the superposition of a large number of fine level nonintersecting terms $(c_\theta \varphi_\theta)$ with a single coarse level term (φ_{θ_0}) :

$$(3.49) \qquad f = \varepsilon \left(\varphi_{[-1,1]} + \sum_{\theta \in \mathcal{M}} \delta \varphi_\theta \right).$$

We set \mathcal{M} as a set of disjoint cells θ from level 2^{2N} with $\theta \subset (-\delta, \delta)$, where $\delta = 2^{-N}$. It is clear that we can choose these cells for \mathcal{M} so that $\#\mathcal{M} = 2^N$. At the central vertex x_θ of each cell θ we have $f(x_\theta) > \varepsilon(1 - \delta) + \delta\varepsilon = \varepsilon$. The “push-the-error algorithm” will produce an inefficient approximation, since it will not select the coarse first term in (3.49) as one might hope. Instead, no such element will be chosen at the coarsest level, and the error will be pushed. At each successive stage the coefficients of the rewritten descendant Courant elements for θ_0 will all again lie beneath the threshold and be further rewritten until all cells are on level 2^{2N} . At that stage they will be combined with the remaining terms in (3.49). The corresponding cells will now have coefficients that exceed the threshold and must be selected, producing at least 2^N terms in the approximant. As indicated above, a desirable algorithm should have anticipated the trap of many small, finely supported elements that may come at a late stage, and would have chosen for this function the approximation (with threshold ε) that consists of a single element, namely $\varepsilon \varphi_{[-1,1]}$.

Trap 3. The final example is one that outmaneuvers a quick remedy to Trap 2, i.e., merely thresholding all small terms at the finest level. For a given $\varepsilon > 0$, we define

$$f = \varepsilon \varphi_{[-1,1]} + \sum_{j=1}^N \delta_j \varphi_{[0,2^{-m_j}]} + \varepsilon \varphi_{[0,2^{-M}]},$$

where $m_j = j^2$, $\delta_j = 2^{-j}\varepsilon$, and $M = 2^{N^2}$. In this example, elements are again building near the origin, but now appear at many levels with small amplitudes. The “push-the-error” algorithm will again take no elements at the coarsest level and push the error to the next level. Continuing with the algorithm, we are forced to take essentially all terms as the approximation to the given function when, optimally, only two terms need be taken.

It is obvious that we can take each of these template examples as building blocks and build functions to cause these problems for all ε , at all locations and scales.

3.3.2. “Push the error” algorithm in the uniform norm ($p = \infty$). In this section we indicate the refinements needed in order to guarantee that the “push the error” algorithm will achieve optimal rates of approximation. As with the “trim and cut” algorithm, we break it down into manageable steps.

• **Description of the algorithm.**

Step 1 (Decompose). For $f \in C(E)$ initially represented by (3.1), we may assume, without loss of generality, that there exists a finest level Θ_J ($J > 0$) such that f is written as

$$(3.50) \qquad f = \sum_{j=0}^J \sum_{\theta \in \Theta_j} b_\theta \varphi_\theta.$$

Step 2 (“*Prune the shrubs*”). In the current algorithm we are not able to organize the cells of Θ into trees as we did in the “trim and cut” method, since, once we rewrite the error on a finer level, adjacent trees are immediately affected and we lose the benefit of their intended organization properties. This step of our algorithm, however, is analogous to Step 3 of the “trim and cut” algorithm. We fix $\varepsilon > 0$ and let $\varepsilon^* := \varepsilon/2$. Our goal is, by discarding small insignificant terms $b_\theta \varphi_\theta$ in the representation of f from (3.50), to prevent our refined algorithm from being trapped by a situation such as that described in “Trap 2” (see the *naive* “push the error” algorithm of §3.3.1). We shall remove such terms, but insure that the resulting uniform error is at most ε^* and denote by Γ the set of all retained cells. In addition, we shall construct a set $\Gamma_f \subset \Gamma$, consisting of “final cells” in Γ .

First, we need to introduce an organizational concept as a replacement for the tree structures of §3.2. We shall say (figuratively) that a cell $\theta \in \Theta$ *sits* on another cell $\theta^\diamond \in \Theta$, if θ is at least as fine as θ^\diamond and its interior (denoted by θ°) intersects the interior of θ^\diamond . Furthermore, for $\theta^\diamond \in \Theta$, we denote the collection of all cells that sit on θ^\diamond by

$$(3.51) \quad \mathcal{Y}_{\theta^\diamond} := \{\theta \in \Theta : \theta^\circ \cap \theta^\diamond \neq \emptyset \text{ and } \text{level}(\theta) \geq \text{level}(\theta^\diamond)\}.$$

The procedure of Step 2 will begin at the finest level and proceed to the coarsest, level by level, constructing sets Γ_f and Γ . To initialize the procedure we put into Γ_f all *significant* cells $\theta \in \Theta_J$, i.e., such that $|b_\theta| > \varepsilon^*$. We place in Γ any cell from Θ_J that sits on a cell from Γ_f .

The inductive step proceeds as follows. Suppose that all cells from Θ_j with levels $j > m$ ($0 \leq m < J$) have already been processed. We now describe how to process Θ_m . We place into Γ_f all cells $\theta^\diamond \in \Theta_m$ that satisfy

$$(3.52) \quad \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta| > \varepsilon^*,$$

and for which there is no $\theta \in \Gamma_f$ from a higher level (i.e., $> m$) that sits on θ^\diamond . A cell θ^\diamond from Θ_m is placed in Γ if there is a cell θ in the current Γ_f that sits on θ^\diamond . We may consider the current version of Γ_f as an intermediate (m -th) version of a final set for Γ . Obviously, a cell θ^\diamond from Θ_m is discarded and not placed in Γ if

$$(3.53) \quad \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta| \leq \varepsilon^*,$$

and there is no $\theta \in \Gamma_f$ from level m or finer that sits on θ^\diamond .

The procedure is terminated after Θ_0 is processed and Step 2 of the algorithm is completed.

The two sets of cells Γ and Γ_f ($\Gamma_f \subset \Gamma \subset \Theta$) produced by Step 2 have the following properties, which follow directly from their construction:

- (i) if $\theta_1, \theta_2 \in \Gamma_f$ and $\text{level}(\theta_1) \neq \text{level}(\theta_2)$, then $\theta_1^\circ \cap \theta_2^\circ = \emptyset$;
- (ii) for each $\theta^\diamond \in \Gamma_f$, the inequality (3.52) holds;
- (iii) for each $\theta^\diamond \in \Gamma$, there exists $\theta \in \Gamma_f$ that sits on θ^\diamond .

We set $f_\Gamma := \sum_{\theta \in \Gamma} b_\theta \varphi_\theta$ and define

$$(3.54) \quad a_\theta := \begin{cases} b_\theta, & \text{if } \theta \in \Gamma, \\ 0, & \text{if } \theta \in \Theta \setminus \Gamma; \end{cases}$$

then obviously

$$(3.55) \quad f_{\Gamma} = \sum_{\theta \in \Theta} a_{\theta} \varphi_{\theta}.$$

It follows from the construction that

$$(3.56) \quad \|f - f_{\Gamma}\|_{\infty} \leq \varepsilon^*.$$

Indeed, to see that this estimate holds, we let \mathcal{D} denote the set of all cells $\theta \in \Theta$ that were discarded during the implementation of Step 2, i.e., $\mathcal{D} = \Theta \setminus \Gamma$. Let $x \in E$ be arbitrary. If $x \notin \bigcup_{\theta \in \mathcal{D}} \theta$, then x does not belong to any cell that was discarded, and so $f_{\Gamma}(x) = f(x)$. On the other hand, if $x \in \bigcup_{\theta \in \mathcal{D}} \theta$, then there exists a cell $\theta^{\diamond} \in \mathcal{D}$ that contains x and has coarsest level. Since θ^{\diamond} was discarded, the inequality (3.53) must hold. It follows that

$$|f(x) - f_{\Gamma}(x)| = \left| \sum_{\theta \in \mathcal{D}} b_{\theta} \varphi_{\theta}(x) \right| \leq \sum_{\theta \in \mathcal{V}_{\theta^{\diamond}}} |b_{\theta}| \leq \varepsilon^*,$$

where we have normalized our elements so that $\|\varphi_{\theta}\|_{\infty} = 1$. This verifies the desired inequality (3.56).

Step 3 (*Push the error*). We now process cells of f_{Γ} with ε^* , starting from the coarsest level Θ_0 and continuing to finer levels. The outcome of this step will be an approximant $\mathcal{A} := \mathcal{A}_{\varepsilon}^P(f)$ of the form

$$(3.57) \quad \mathcal{A} = \sum_{j=0}^J \mathcal{A}_j := \sum_{j=0}^J \sum_{\theta \in \Lambda_j} d_{\theta} \varphi_{\theta},$$

where $\Lambda_j \subset \Theta_j$ and Λ_j will depend on f .

We use the notation

$$\mathcal{X}_{\theta^{\diamond}} := \{\theta \in \Theta : \theta^{\circ} \cap \theta^{\diamond} \neq \emptyset \text{ and } \text{level}(\theta) = \text{level}(\theta^{\diamond})\}$$

for cells from the same level as θ^{\diamond} which are adjacent to it.

We start from the representation of f_{Γ} in (3.55). We define $\tilde{\Lambda}_0$ as the set of all $\theta \in \Theta_0$ such that $|a_{\theta}| > \varepsilon^*$ ($\|\varphi_{\theta}\|_{\infty} = 1$), and we set $\Lambda_0 := \bigcup_{\theta \in \tilde{\Lambda}_0} \mathcal{X}_{\theta}$. We denote

$$\mathcal{A}_0 := \sum_{\theta \in \Lambda_0} a_{\theta} \varphi_{\theta} =: \sum_{\theta \in \Lambda_0} d_{\theta} \varphi_{\theta}.$$

For each $\theta^{\diamond} \in \Theta_j$, $\varphi_{\theta^{\diamond}}$ can be represented as a linear combination of φ_{θ} 's with $\theta \in \Theta_{j+1}$. We use this to rewrite (represent) all remaining terms $a_{\theta} \varphi_{\theta}$, $\theta \in \Theta_0 \setminus \Lambda_0$, at the next level and add the resulting terms to the corresponding terms $a_{\theta} \varphi_{\theta}$, $\theta \in \Theta_1$. We denote by $d_{\theta} \varphi_{\theta}$, $\theta \in \Theta_1$, the new terms, and therefore obtain a representation of f in the form

$$f = \mathcal{A}_0 + \sum_{\theta \in \Theta_1} d_{\theta} \varphi_{\theta} + \sum_{j=2}^J \sum_{\theta \in \Theta_j} a_{\theta} \varphi_{\theta}.$$

Continuing with the next level, we define $\tilde{\Lambda}_1$ as the set of all $\theta \in \Theta_1$ such that $|d_{\theta}| > \varepsilon^*$, set $\Lambda_1 := \bigcup_{\theta \in \tilde{\Lambda}_1} \mathcal{X}_{\theta}$, and define $\mathcal{A}_1 := \sum_{\theta \in \Lambda_1} d_{\theta} \varphi_{\theta}$. As for the previous level, we rewrite the remaining terms $d_{\theta} \varphi_{\theta}$, $\theta \in \Theta_1 \setminus \Lambda_1$, at the next level and

add the resulting terms to the corresponding terms $a_\theta \varphi_\theta$, $\theta \in \Theta_2$. We obtain the following representation of f :

$$f = \mathcal{A}_0 + \mathcal{A}_1 + \sum_{\theta \in \Theta_2} d_\theta \varphi_\theta + \sum_{j=3}^J \sum_{\theta \in \Theta_j} a_\theta \varphi_\theta.$$

We continue in this way until we reach the highest level of cells Θ_J . At level Θ_J , we define $\tilde{\Lambda}_J$, Λ_J , and \mathcal{A}_J as above and discard all terms $d_\theta \varphi_\theta$, $\theta \in \Theta_J \setminus \Lambda_J$. We finally obtain our approximant $\mathcal{A} = \mathcal{A}_\varepsilon^P(f)$ in the form (3.57). We denote $\Lambda := \Lambda_\varepsilon := \bigcup_{j=0}^J \Lambda_j$ and $\tilde{\Lambda} := \tilde{\Lambda}_\varepsilon := \bigcup_{j=0}^J \tilde{\Lambda}_j$, and so $\mathcal{A} = \sum_{\theta \in \Lambda} d_\theta \varphi_\theta$.

Since we throw away only elements $d_\theta \varphi_\theta$ with $|d_\theta| \leq \varepsilon^*$ at the finest level Θ_J , we have the estimate

$$\|f - \mathcal{A}\|_\infty \leq \left\| \sum_{\theta \in \Theta_J \setminus \Lambda_J} d_\theta \varphi_\theta \right\|_\infty \leq \varepsilon^*,$$

and hence, using (3.56),

$$(3.58) \quad \|f - \mathcal{A}\|_\infty \leq 2\varepsilon^* = \varepsilon.$$

This completes Step 3 and with that the description of the algorithm.

We want to point out an important distinction between the “push the error” steps in the above algorithm and the “naive” algorithm described in §3.3.1. The difference is that each time we put a significant term $d_\theta \varphi_\theta$ ($|d_\theta| > \varepsilon^*$) into \mathcal{A} we also include the neighboring terms (i.e., from the index collection \mathcal{X}_θ). This prevents our algorithm from being defeated by a situation like that described in “Trap 3” in §3.3.1.

• **Error estimation for the “push the error” algorithm.** Suppose “push the error” is applied to a function f with $\varepsilon > 0$, and $\mathcal{A}_\varepsilon^P(f)$ is the approximant obtained: $\mathcal{A}_\varepsilon^P(f) := \sum_{\theta \in \Lambda_\varepsilon} d_\theta \varphi_\theta$. As in the “trim and cut” method, we use the corresponding notation

$$n(\varepsilon) := \#\Lambda_\varepsilon, \quad \mathbb{A}_{n(\varepsilon)}^P(f)_\infty := \mathbb{A}_{n(\varepsilon)}^P(f, \mathcal{T})_\infty := \|f - \mathcal{A}_\varepsilon^P(f)\|_\infty,$$

and

$$\mathbb{A}_n^P(f)_\infty := \mathbb{A}_n^P(f, \mathcal{T})_\infty := \inf\{\mathbb{A}_{n(\varepsilon)}^P(f)_\infty : n(\varepsilon) \leq n\}.$$

We remark that if $f \in B_\tau^\alpha(\mathcal{T})$, then by the Embedding Theorem 2.7 it follows that f is continuous. Estimates (3.59) and (3.60), established in the following theorem, imply uniform convergence of the “push the error” approximants to f and provide the necessary rates of approximation by the method.

Theorem 3.11. *If $f \in B_\tau^\alpha(\mathcal{T})$, $\alpha \geq 1$, $\tau := 1/\alpha$, then for each $\varepsilon > 0$,*

$$(3.59) \quad \mathbb{A}_{n(\varepsilon)}^P(f)_\infty \leq \varepsilon \quad \text{and} \quad n(\varepsilon) \leq c\varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau,$$

where $c = 6N_0^3$. Furthermore, we have

$$(3.60) \quad \mathbb{A}_n^P(f)_\infty \leq cn^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})}, \quad n = 1, 2, \dots$$

with $c = (6N_0^3)^\alpha$.

Proof. In order to prove (3.59), we first observe that the direct approximation estimate $\mathbb{A}_{n(\varepsilon)}^P(f)_\infty \leq \varepsilon$ follows from inequality (3.58) in the construction of the algorithm. Therefore it only remains to show that $\#\Lambda_\varepsilon \leq c\varepsilon^{-\tau}\|f\|_{B_\tau^\infty(\mathcal{T})}^\tau$. Clearly,

$$(3.61) \quad \#\Lambda_\varepsilon \leq (N_0 + 1)(\#\tilde{\Lambda}_\varepsilon),$$

and we need only estimate the cardinality of $\tilde{\Lambda} := \tilde{\Lambda}_\varepsilon$. We split $\tilde{\Lambda}$ into two disjoint sets, $\tilde{\Lambda}_f$ and $\tilde{\Lambda}_r$. We define $\tilde{\Lambda}_f$ as the set of all final cells in $\tilde{\Lambda}$, that is, the set of all $\theta \in \tilde{\Lambda}$ for which there is no $\theta' \in \tilde{\Lambda}$ of a higher level sitting on θ . We set $\tilde{\Lambda}_r := \tilde{\Lambda} \setminus \tilde{\Lambda}_f$.

We shall make repeated use of the following simple lemma.

Lemma 3.12. *Suppose $\mathcal{M} \subset \Theta$ satisfies the condition that cells from different levels do not have interiors that intersect. Then each $\theta \in \Theta$ may sit on at most $N_0 + 1$ cells from \mathcal{M} .*

Proof. The simple hypothesis of the lemma just states that for a cell θ_2 to sit on a cell θ_1 , it must be on the same level; but there can be at most $N_0 + 1$ such cells. \square

We first estimate the number of elements $\#\Gamma_f$ that arise as final cells in Step 2. For each $\theta^\diamond \in \Gamma_f$, we have, by (3.52),

$$(3.62) \quad \varepsilon^* < \sum_{\theta \in \mathcal{V}_{\theta^\diamond}} |b_\theta| \leq \left(\sum_{\theta \in \mathcal{V}_{\theta^\diamond}} |b_\theta|^\tau \right)^{1/\tau} \quad (\tau \leq 1).$$

Clearly, Γ_f satisfies the hypothesis of Lemma 3.12 (see Property (i) of Γ_f , which is stated following (3.52)), and hence each $\theta \in \Theta$ may sit on at most $N_0 + 1$ cells from Γ_f . Using this together with (3.62), we obtain

$$\|f\|_{B_\tau^\infty(\mathcal{T})}^\tau := \sum_{\theta \in \Theta} |b_\theta|^\tau \geq (N_0 + 1)^{-1} \sum_{\theta^\diamond \in \Gamma_f} \sum_{\theta \in \mathcal{V}_{\theta^\diamond}} |b_\theta|^\tau \geq (N_0 + 1)^{-1} (\#\Gamma_f) (\varepsilon^*)^\tau,$$

which, since $\tau \leq 1$, implies

$$(3.63) \quad \#\Gamma_f \leq 2(N_0 + 1)\varepsilon^{-\tau}\|f\|_{B_\tau^\infty(\mathcal{T})}^\tau.$$

We next estimate $\#\tilde{\Lambda}_f$, the number of final cells for the index set $\tilde{\Lambda}$ constructed in Step 3. Clearly from that construction, a cell $\theta \in \tilde{\Lambda}$ may occur only if $\theta \in \Gamma$, and hence $\tilde{\Lambda} \subset \Gamma$. On the other hand, from Step 2, for each $\theta \in \Gamma$ there exists $\theta' \in \Gamma_f$ sitting on θ . Therefore, for each $\theta \in \tilde{\Lambda}_f$ there exists $\theta' \in \Gamma_f$ sitting on θ . But $\tilde{\Lambda}_f$ satisfies the hypothesis of Lemma 3.12 (with \mathcal{M} replaced by $\tilde{\Lambda}_f$), and hence a cell $\theta \in \Gamma_f$ may sit on at most $N_0 + 1$ cells from $\tilde{\Lambda}_f$. From this and (3.63), we have

$$(3.64) \quad \#\tilde{\Lambda}_f \leq (N_0 + 1)(\#\Gamma_f) \leq 2(N_0 + 1)^2\varepsilon^{-\tau}\|f\|_{B_\tau^\infty(\mathcal{T})}^\tau.$$

To complete the estimate for $\#\tilde{\Lambda}$, we must estimate $\#\tilde{\Lambda}_r$. Suppose $\theta^\diamond \in \tilde{\Lambda}_r := \tilde{\Lambda} \setminus \tilde{\Lambda}_f$, and let $\theta' \in \tilde{\Lambda}$ be a cell sitting on θ^\diamond with $\text{level}(\theta') > \text{level}(\theta^\diamond)$ and such that $\text{level}(\theta')$ is the minimum of the levels of all cells in $\tilde{\Lambda}$ sitting on θ^\diamond . Such a cell exists, by the definition of $\tilde{\Lambda}_r$, but it is possibly not unique. We denote by $\mathcal{Z}_{\theta^\diamond}$ the set of all $\theta \in \Gamma$ which, while “pushing the error” from θ^\diamond in Step 3, have contributed to the term $d_{\theta'}\varphi_{\theta'}$. Due to the minimality of θ' , we see that

$$(3.65) \quad d_{\theta'} = d_{\theta'}\varphi_{\theta'}(v_{\theta'}) = \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} b_\theta\varphi_\theta(v_{\theta'}),$$

where $v_{\theta'}$ is the “central vertex” of θ' . Since $\theta' \in \tilde{\Lambda}$, then $|d_{\theta'}| > \varepsilon^*$, and hence, using (3.65),

$$(3.66) \quad \varepsilon^* < |d_{\theta'}| \leq \left\| \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} b_\theta \varphi_\theta \right\|_\infty \leq \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta| \leq \left(\sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta|^\tau \right)^{1/\tau} \quad (\tau \leq 1).$$

It is easily seen that each $\theta \in \mathcal{Z}_{\theta^\diamond}$ satisfies the following properties:

- (a) $\theta \supset \theta'$,
- (b) $\text{level}(\theta^\diamond) < \text{level}(\theta) \leq \text{level}(\theta')$,
- (c) the “central vertex” of θ lies on θ^\diamond , and hence θ sits on θ^\diamond .

Property (a) follows by observing that the support of an element which is rewritten at a finer level always contains the supports of the contributing finer elements. Property (b) holds, since $\mathcal{X}_{\theta^\diamond} \subset \Lambda$, and hence no terms $b_\theta \varphi_\theta$ with $\text{level}(\theta) \leq \text{level}(\theta^\diamond)$ may contribute to $d_{\theta'} \varphi_{\theta'}$. Note that it is possible that there are θ that satisfy properties (a)-(c) above but do not belong to $\mathcal{Z}_{\theta^\diamond}$.

Next, we show that each $\theta \in \Gamma$ may belong to at most $N_0 + 1$ sets \mathcal{Z}_{θ^*} with $\theta^* \in \tilde{\Lambda}_r$. Indeed, let $\theta \in \Gamma$ and suppose $\theta^\diamond \in \tilde{\Lambda}_r$ is such that $\theta \in \mathcal{Z}_{\theta^\diamond}$. In the following, we shall use the notation from above that involves θ^\diamond , but we will consider such θ^\diamond as arbitrary in $\tilde{\Lambda}$. Let \mathcal{M}_θ denote the set of all $\theta^\sharp \in \tilde{\Lambda}$ such that $\theta \in \mathcal{Z}_{\theta^\sharp}$. In particular, $\theta^\diamond \in \mathcal{M}_\theta$. We fix \mathcal{M}_θ and show that it satisfies the hypothesis of Lemma 3.12. Indeed, let $\theta_1, \theta_2 \in \mathcal{M}_\theta$ from different levels. But this implies $\theta \in \mathcal{Z}_{\theta_j}$ ($j = 1, 2$), and we may as well consider $\theta_1 = \theta^\diamond$ and say $\theta_2 = \theta^\sharp$, where $\text{level}(\theta^\sharp) \neq \text{level}(\theta^\diamond)$. Evidently, $\text{level}(\theta^\sharp) < \text{level}(\theta')$, from property (b) applied to θ^\sharp and θ .

By symmetry, we may assume $\text{level}(\theta^\sharp) < \text{level}(\theta^\diamond)$. If $(\theta^\sharp)^\diamond \cap (\theta^\diamond)^\diamond \neq \emptyset$, then θ^\diamond sits on θ^\sharp and hence, since $\text{level}(\theta) > \text{level}(\theta^\diamond)$, θ cannot be in $\mathcal{Z}_{\theta^\sharp}$, which is a contradiction. Therefore, $(\theta^\sharp)^\diamond \cap (\theta^\diamond)^\diamond = \emptyset$, which verifies the hypothesis of Lemma 3.12.

Now that Lemma 3.12 can be applied to \mathcal{M}_θ , then θ (as any other cell from Θ) may sit on at most $N_0 + 1$ cells $\theta^* \in \mathcal{M}_\theta$. Therefore, θ may belong to at most $N_0 + 1$ such sets \mathcal{Z}_{θ^*} with $\theta^* \in \tilde{\Lambda}_r$. Using this and (3.66), we obtain

$$\|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau \geq \sum_{\theta \in \Gamma} |b_\theta|^\tau \geq (N_0 + 1)^{-1} \sum_{\theta^\diamond \in \tilde{\Lambda}_r} \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta|^\tau \geq (N_0 + 1)^{-1} (\#\tilde{\Lambda}_r) (\varepsilon^*)^\tau.$$

Therefore, it follows (recall that $\tau < 1$) that

$$\#\tilde{\Lambda}_r \leq 2(N_0 + 1) \varepsilon^{-\tau} \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau.$$

We combine this estimate with (3.61) and (3.64) to obtain the desired estimate of $\#\Lambda_\varepsilon$ in (3.59). Estimate (3.60) follows immediately from (3.59). \square

The following lemma will be needed in §5.

Lemma 3.13. *Let $f = f^0 + f^1$, where $f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta$, $f^j = \sum_{\theta \in \Theta} b_\theta^j \varphi_\theta$ ($j = 0, 1$), and $b_\theta = b_\theta^0 + b_\theta^1$ (all $\theta \in \Theta$), and suppose*

$$\mathcal{N}_j := \left(\sum_{\theta \in \Theta} |b_\theta^j|^{\tau_j} \right)^{1/\tau_j} < \infty \quad (j = 0, 1),$$

where $\alpha_0, \alpha_1 \geq 1$ and $\tau_0 := 1/\alpha_0$, $\tau_1 := 1/\alpha_1$. Furthermore, suppose that “push the error” is applied using the above representation of f , with $\varepsilon := \varepsilon_0 + \varepsilon_1$, where

$\varepsilon_0, \varepsilon_1 > 0$. Then we have

$$(3.67) \quad \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^P(f)_\infty \leq \varepsilon_0 + \varepsilon_1,$$

$$(3.68) \quad n(\varepsilon_0 + \varepsilon_1) \leq c\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + c\varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1},$$

where $c = 6N_0^3$. Consequently,

$$(3.69) \quad \mathbb{A}_n^P(f)_\infty \leq c_0 n^{-\alpha_0} \mathcal{N}_0 + c_1 n^{-\alpha_1} \mathcal{N}_1, \quad n = 1, 2, \dots,$$

with $c_j = (12N_0^3)^{\alpha_j}$.

Proof. We follow in the footsteps of the proof of Theorem 3.11. We shall use the notation from there, and only indicate the differences as they arise. We denote $\varepsilon^* := \varepsilon_0^* + \varepsilon_1^*$ with $\varepsilon_j^* := \varepsilon_j/2$, $j = 0, 1$. Estimate (3.67) is immediate from the description of the algorithm.

It remains to provide estimate (3.68) for the number of terms used in the approximation. As in (3.61), we have

$$(3.70) \quad n(\varepsilon_0 + \varepsilon_1) := \#\Lambda_\varepsilon \leq (N_0 + 1)(\#\tilde{\Lambda}_\varepsilon),$$

where we denote $\tilde{\Lambda} := \tilde{\Lambda}_\varepsilon$, and $\tilde{\Lambda}_f$ and $\tilde{\Lambda}_r$ have the same definitions, proceeding exactly as in the proof of Theorem 3.11. Continuing as there, we have to estimate $\#\Gamma_f$. For each $\theta^\diamond \in \Gamma_f$, we have, by (3.52) and the fact that $0 < \tau_j \leq 1$ ($j = 0, 1$), that

$$\begin{aligned} \varepsilon_0^* + \varepsilon_1^* = \varepsilon^* &< \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta| \leq \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^0| + \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^1| \\ &\leq \left(\sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^0|^{\tau_0} \right)^{1/\tau_0} + \left(\sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^1|^{\tau_1} \right)^{1/\tau_1}. \end{aligned}$$

From this, it follows that, for each $\theta^\diamond \in \Gamma_f$, at least one of

$$(3.71) \quad \varepsilon_0^* < \left(\sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^0|^{\tau_0} \right)^{1/\tau_0} \quad \text{or} \quad \varepsilon_1^* < \left(\sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^1|^{\tau_1} \right)^{1/\tau_1}$$

must hold. We denote by Γ_f^0 and Γ_f^1 the sets of all $\theta^\diamond \in \Gamma_f$ such that the respective condition from (3.71) holds for either $j = 0$ or $j = 1$. For $j = 0, 1$, we have similarly, as in the proof of Theorem 3.11,

$$\mathcal{N}_j^{\tau_j} := \sum_{\theta \in \Theta} |b_\theta^j|^{\tau_j} \geq (N_0 + 1)^{-1} \sum_{\theta^\diamond \in \Gamma_f^j} \sum_{\theta \in \mathcal{Y}_{\theta^\diamond}} |b_\theta^j|^{\tau_j} \geq (N_0 + 1)^{-1} (\#\Gamma_f^j) (\varepsilon_j^*)^{\tau_j},$$

and hence ($\tau_j \leq 1$)

$$\#\Gamma_f^j \leq 2(N_0 + 1) \varepsilon_j^{-\tau_j} \mathcal{N}_j^{\tau_j}.$$

Therefore,

$$\begin{aligned} \#\tilde{\Lambda}_f &\leq (N_0 + 1)(\#\Gamma_f) \leq (N_0 + 1)(\#\Gamma_f^0 + \#\Gamma_f^1) \\ (3.72) \quad &\leq 2(N_0 + 1)^2 \left(\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + \varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} \right). \end{aligned}$$

To complete the proof, we must next estimate $\#\tilde{\Lambda}_r$. For each $\theta^\diamond \in \tilde{\Lambda}_r$, we define $\theta' \in \tilde{\Lambda}$ and $\mathcal{Z}_{\theta^\diamond}$ exactly as in the proof of Theorem 3.11. Similarly as in (3.66), we

have

$$\begin{aligned} \varepsilon_0^* + \varepsilon_1^* = \varepsilon^* &< \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta| \leq \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^0| + \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^1| \\ &\leq \left(\sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^0|^{\tau_0} \right)^{1/\tau_0} + \left(\sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^1|^{\tau_1} \right)^{1/\tau_1}. \end{aligned}$$

From this, it follows that, for each $\theta^\diamond \in \tilde{\Lambda}_r$, at least one of

$$(3.73) \quad \varepsilon_0^* < \left(\sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^0|^{\tau_0} \right)^{1/\tau_0}$$

or

$$(3.74) \quad \varepsilon_1^* < \left(\sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^1|^{\tau_1} \right)^{1/\tau_1}$$

must hold. We denote by $\tilde{\Lambda}_r^0$ and $\tilde{\Lambda}_r^1$ the sets of all $\theta^\diamond \in \tilde{\Lambda}_r$ for which (3.73) and (3.74) hold, respectively. As in the proof of Theorem 3.11, each $\theta \in \Theta$ may belong to at most $N_0 + 1$ sets $\mathcal{Z}_{\theta^\diamond}$, $\theta^\diamond \in \tilde{\Lambda}_r$. Therefore, for $j = 0, 1$,

$$\mathcal{N}_j^{\tau_j} \geq \sum_{\theta \in \Gamma} |b_\theta^j|^{\tau_j} \geq (N_0 + 1)^{-1} \sum_{\theta^\diamond \in \tilde{\Lambda}_r^j} \sum_{\theta \in \mathcal{Z}_{\theta^\diamond}} |b_\theta^j|^{\tau_j} \geq (N_0 + 1)^{-1} (\#\tilde{\Lambda}_r^j) (\varepsilon_j^*)^{\tau_j},$$

and hence

$$\#\tilde{\Lambda}_r^j \leq 2(N_0 + 1) \varepsilon_j^{-\tau_j} \mathcal{N}_j^{\tau_j}, \quad j = 0, 1.$$

Therefore,

$$\#\tilde{\Lambda}_r \leq \#\tilde{\Lambda}_r^0 + \#\tilde{\Lambda}_r^1 \leq 2(N_0 + 1) \left(\varepsilon_0^{-\tau_0} \mathcal{N}_0^{\tau_0} + \varepsilon_1^{-\tau_1} \mathcal{N}_1^{\tau_1} \right).$$

This estimate, together with (3.70) and (3.72), implies (3.68) (since $N_0 \geq 3$). Estimate (3.69) follows by using $\varepsilon_j := (2c)^{\alpha_j} n^{-\alpha_j} \mathcal{N}_j$ ($j = 0, 1$) in (3.67) and (3.68) to obtain $n(\varepsilon_0 + \varepsilon_1) \leq n$, and so $\mathbb{A}_n^P(f)_\infty \leq \mathbb{A}_{n(\varepsilon_0 + \varepsilon_1)}^P(f)_\infty \leq \varepsilon_0 + \varepsilon_1$. \square

4. BEST n -TERM COURANT ELEMENT APPROXIMATION

In this section, we assume that \mathcal{T} is a locally regular triangulation of a bounded polygonal domain E with parameters $N_0, M_0, r, \rho, \delta$, and $\#\mathcal{T}_0$ (see §2.1). We denote by $\Phi_{\mathcal{T}}$ the collection of all Courant elements φ_θ generated by \mathcal{T} . Notice that $\Phi_{\mathcal{T}}$ is not a basis; $\Phi_{\mathcal{T}}$ is redundant. We consider nonlinear n -term approximation in $L_p(E)$ ($0 < p \leq \infty$) from $\Phi_{\mathcal{T}}$, where we identify $L_\infty(E)$ as $C(E)$. Our main goal is to characterize the approximation spaces generated by this approximation, with emphasis on the case $p = \infty$. We let $\Sigma_n(\mathcal{T})$ denote the nonlinear set consisting of all continuous piecewise linear functions S of the form

$$S = \sum_{\theta \in \mathcal{M}} a_\theta \varphi_\theta,$$

where $\mathcal{M} \subset \Theta(\mathcal{T})$, $\#\mathcal{M} \leq n$, and \mathcal{M} may vary with S . We denote by $\sigma_n(f, \mathcal{T})_p$ the best L_p -approximation of $f \in L_p(E)$ from $\Sigma_n(\mathcal{T})$:

$$\sigma_n(f, \mathcal{T})_p := \inf_{S \in \Sigma_n(\mathcal{T})} \|f - S\|_p.$$

In order to characterize the approximation spaces generated by $(\sigma_n(f, \mathcal{T})_p)$, we begin in this section by first proving a companion pair of Jackson and Bernstein

inequalities, and then follow with the usual techniques of interpolation of operators (see for example [6], [15], [13]).

In the following, we assume in general that $0 < p \leq \infty$, and that $\alpha \geq 1$ for $p = \infty$ and $\alpha > 0$ if $p < \infty$; in either case we set $1/\tau := \alpha + 1/p$.

Theorem 4.1 (Jackson estimate). *If $f \in B_\tau^\alpha(\mathcal{T})$, then*

$$(4.1) \quad \sigma_n(f, \mathcal{T})_p \leq cn^{-\alpha} \|f\|_{B_\tau^\alpha(\mathcal{T})},$$

where c depends only on α , p and the parameters of \mathcal{T} .

Proof. Estimate (4.1) follows from any of our constructive algorithms as formulated in the corresponding Theorems 3.1, 3.7, 3.9, or 3.11. \square

Theorem 4.2 (Bernstein estimate). *If $S \in \Sigma_n(\mathcal{T})$, then*

$$(4.2) \quad \|S\|_{B_\tau^\alpha(\mathcal{T})} \leq cn^\alpha \|S\|_p,$$

where c depends only on α , p , and the parameters of \mathcal{T} .

Proof. We shall prove estimate (4.2) only in the case $p = \infty$. For the proof when $p < \infty$, see [11]. Suppose $S \in \Sigma_n^k(\mathcal{T})$ and $S =: \sum_{\theta \in \mathcal{M}} c_\theta \varphi_\theta$, where $\mathcal{M} \subset \Theta(\mathcal{T})$ and $\#\mathcal{M} \leq n$. Let Λ be the set of all triangles $\Delta \in \mathcal{T}$ that are involved in all cells $\theta \in \mathcal{M}$. Then $S = \sum_{\Delta \in \Lambda} S_\Delta$, where $S_\Delta =: \mathbb{1}_\Delta \cdot P_\Delta$, P_Δ a linear polynomial. Evidently, $\#\Lambda \leq N_0 \#\mathcal{M} \leq cn$.

We shall utilize the natural tree structure in \mathcal{T} induced by the inclusion relation: Each triangle $\Delta \in \mathcal{T}_m$ has (contains) $\leq M_0$ children in \mathcal{T}_{m+1} and one parent in \mathcal{T}_{m-1} , if $m \geq 1$. Let Γ_0 be the set of all $\Delta \in \mathcal{T}$ such that $\Delta \supset \Delta'$ for some $\Delta' \in \Lambda$. We denote by Γ_b the set of all *branching triangles* in Γ_0 (triangles with more than one child in Γ_0) and by Γ'_b the set of all *children of branching triangles* in \mathcal{T} (which may or may not belong to Γ_0). Now, we extend Γ_0 to $\Gamma := \Gamma_0 \cup \Gamma'_b$. We also extend Λ to $\tilde{\Lambda} := \Lambda \cup \Gamma_b \cup \Gamma'_b$. In addition, we introduce the following subsets of Γ : the set Γ_f of all *final triangles* in Γ (triangles in Γ containing no other triangles in Γ), the set $(\Gamma_0)_f$ of the final triangles in Γ_0 , and the set $\Gamma_{ch} := \Gamma \setminus \tilde{\Lambda}$ of all *chain triangles*. Note that each triangle $\Delta \in \Gamma_{ch}$ has exactly one child in Γ . We may argue as we did for trees of cells in (3.9) that the number of branching triangles does not exceed the number of final triangles, $\#\Gamma_b \leq \#(\Gamma_0)_f$, and since $(\Gamma_0)_f \subset \Lambda$, then $\#\Gamma_b \leq cn$. Using this, we have $\#\Gamma'_b \leq M_0 \#\Gamma_b \leq cn$, $\#\Gamma_f \leq \#\Lambda + \#\Gamma'_b \leq cn$, and $\#\tilde{\Lambda} \leq \#\Lambda + \#\Gamma_b + \#\Gamma'_b \leq cn$. Keep in mind, however, that $\#\Gamma_{ch}$ can be much larger than n .

We next estimate $|S|_{B_\tau^\alpha(\mathcal{T})}^\tau := \sum_{\Delta \in \mathcal{T}} |\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau$, where $\tau := 1/\alpha$ (see (2.5) for the notation). We denote, for $m \geq 0$, $S_m := \sum_{\theta \in \mathcal{M}, \text{level}(\theta) \leq m} c_\theta \varphi_\theta$. We shall use that, for $\Delta \in \mathcal{T}_m$,

$$(4.3) \quad \mathbb{S}_\Delta(S)_\tau = \mathbb{S}_\Delta(S - S_m)_\tau \leq \|S - S_m\|_{L_\tau(\Omega_\Delta)}$$

and, also, $\mathbb{S}_\Delta(S)_\tau \leq \|S\|_{L_\tau(\Omega_\Delta)}$. Recall that Ω_Δ is the union of the collection of all triangles from the same level as Δ and which share a vertex. We denote

$$\mathcal{H}_m := \{\Delta \in \mathcal{T}_m : \Delta \subset \Omega_{\Delta'}, \text{ for some } \Delta' \in \tilde{\Lambda} \cap \mathcal{T}_m\} \quad \text{and} \quad \mathcal{H} := \bigcup_{m \geq 0} \mathcal{H}_m.$$

Evidently, $\#\mathcal{H}_m \leq 3N_0 \#\tilde{\Lambda} \leq cn$ (the valence of each vertex is $\leq N_0$). We consider two possibilities for each $\Delta \in \mathcal{T}$: (a) $\Delta \in \mathcal{H}$, or (b) $\Delta \in \mathcal{T} \setminus \mathcal{H}$:

(a) If $\Delta \in \mathcal{H}_m$, then $\Omega_\Delta \supset \Delta'$ for some $\Delta' \in \tilde{\Lambda} \cap \mathcal{T}_m$. Using (2.3), we obtain

$$|\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau \leq |\Delta|^{-1} \|S\|_{L_\tau(\Omega_\Delta)}^\tau \leq |\Delta|^{-1} |\Omega_\Delta| \|S\|_\infty^\tau \leq c \|S\|_\infty^\tau.$$

Therefore, by summing over all $m \geq 0$, we obtain in this case

$$\begin{aligned} \sum_{\Delta \in \mathcal{H}} |\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau &= \sum_{m \geq 0} \sum_{\Delta \in \mathcal{H}_m} |\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau \\ &\leq c \|S\|_\infty^\tau \sum_{m \geq 0} \#\mathcal{H}_m \\ (4.4) \qquad &= c \|S\|_\infty^\tau \#\mathcal{H} \leq cn \|S\|_\infty^\tau. \end{aligned}$$

(b) Let $\Delta \in \mathcal{T}_m \setminus \mathcal{H}_m$. Then $\Omega_\Delta =: \bigcup_{j=1}^{n_\Delta} \Delta_j$ for some $\Delta_j \in (\Gamma_{ch} \cap \mathcal{T}_m) \cup (\mathcal{T}_m \setminus \Gamma)$, $j = 1, \dots, n_\Delta$, with $n_\Delta \leq 3N_0$. We have, using (4.3),

$$\mathbb{S}_\Delta(S)_\tau^\tau = \mathbb{S}_\Delta(S - S_m)_\tau^\tau \leq \sum_{j=1}^{n_\Delta} \|S - S_m\|_{L_\tau(\Delta_j)}^\tau.$$

Note that if $\Delta_j \in \mathcal{T}_m \setminus \Gamma$, then $S|_{\Delta_j} = S_m|_{\Delta_j}$ and hence $\|S - S_m\|_{L_\tau(\Delta_j)} = 0$. Suppose $\Delta_j \in \Gamma_{ch} \cap \mathcal{T}_m$. For each $\Delta \in \Gamma_{ch}$, we shall denote by $\tilde{\Delta}$ ($\tilde{\Delta} \neq \Delta$) the unique largest triangle of $\tilde{\Lambda}$ contained in Δ . Clearly, we have $S|_{\Delta_j \setminus \tilde{\Delta}_j} = S_m|_{\Delta_j \setminus \tilde{\Delta}_j} = \mathbb{1}_{\Delta_j \setminus \tilde{\Delta}_j} \cdot P_{\Delta_j}$ and $S_m|_{\Delta_j} = \mathbb{1}_{\Delta_j} \cdot P_{\Delta_j}$, where P_{Δ_j} is a linear polynomial. Therefore,

$$\begin{aligned} \|S - S_m\|_{L_\tau(\Delta_j)}^\tau &= \|S - S_m\|_{L_\tau(\tilde{\Delta}_j)}^\tau \\ &\leq c |\tilde{\Delta}_j| (\|S\|_\infty^\tau + \|P_{\Delta_j}\|_{L_\infty(\tilde{\Delta}_j)}^\tau) \leq c |\tilde{\Delta}_j| \|S\|_\infty^\tau, \end{aligned}$$

where we used that $\|P_{\Delta_j}\|_{L_\infty(\tilde{\Delta}_j)} \leq \|P_{\Delta_j}\|_{L_\infty(\Delta_j)} \leq c \|P_{\Delta_j}\|_{L_\infty(\Delta_j \setminus \tilde{\Delta}_j)} \leq c \|S\|_\infty$, applying Lemma 2.1. From the above, it follows that

$$|\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau \leq c \|S\|_\infty^\tau \sum_{1 \leq j \leq n_\Delta, \Delta_j \in \Gamma_{ch} \cap \mathcal{T}_m} |\tilde{\Delta}_j|/|\Delta_j|$$

and hence

$$\sum_{\Delta \in \mathcal{T}_m \setminus \mathcal{H}_m} |\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau \leq c \|S\|_\infty^\tau \sum_{\Delta \in \Gamma_{ch} \cap \mathcal{T}_m} |\tilde{\Delta}|/|\Delta|.$$

Summing over $m \geq 0$ in this case as well, we find that

$$\begin{aligned} \sum_{\Delta \in \mathcal{T} \setminus \mathcal{H}} |\Delta|^{-1} \mathbb{S}_\Delta(S)_\tau^\tau &\leq c \|S\|_\infty^\tau \sum_{\Delta \in \Gamma_{ch}} |\tilde{\Delta}|/|\Delta| \\ &\leq c \|S\|_\infty^\tau \sum_{\Delta' \in \tilde{\Lambda}} \sum_{\Delta \in \mathcal{T}, \Delta \supset \Delta'} |\Delta'|/|\Delta| \\ (4.5) \qquad &\leq c \|S\|_\infty^\tau \sum_{\Delta' \in \tilde{\Lambda}} \sum_{j=0}^{\infty} \rho^j \leq c \|S\|_\infty^\tau \#\tilde{\Lambda} \leq cn \|S\|_\infty^\tau, \end{aligned}$$

where we have once switched the order of summation and used that $|\Delta'| \leq \rho |\Delta|$ if Δ' is a child of Δ (see (2.2)).

Combining inequalities (4.4) and (4.5), we obtain $|S|_{B_\tau^\alpha(\mathcal{T})}^\tau \leq cn \|S\|_\infty^\tau$, which is equivalent to (4.2). \square

We define the approximation space $A_q^\gamma(L_p) := A_q^\gamma(L_p, \mathcal{T})$ generated by the n -term Courant element approximation to be the set of all functions $f \in L_p(E)$ such that

$$(4.6) \quad \|f\|_{A_q^\gamma(L_p)} := \|f\|_p + \left(\sum_{n=1}^{\infty} (n^\gamma \sigma_n(f, \mathcal{T})_p)^q \frac{1}{n} \right)^{1/q} < \infty,$$

with the usual modification when $q = \infty$.

For a fixed LR-triangulation \mathcal{T} , we denote by $K(f, t) := K(f, t; L_p, B_\tau^\alpha(\mathcal{T}))$ the K -functional as defined in (2.30). The Jackson and Bernstein estimates from Theorem 4.1 and Theorem 4.2 yield (see, e.g., Theorem 3.16 of [15] and its proof) the following direct and inverse estimates:

$$(4.7) \quad \sigma_n(f, \mathcal{T})_p \leq cK(f, n^{-\alpha})$$

and

$$(4.8) \quad K(f, n^{-\alpha}) \leq cn^{-\alpha} \left(\|f\|_p + \left(\sum_{k=1}^n \frac{1}{k} (k^\alpha \sigma_k(f, \mathcal{T})_p)^{p^*} \right)^{1/p^*} \right), \quad p^* := \min\{p, 1\},$$

where c depends only on α , p , and the parameters of \mathcal{T} .

The following characterization of the approximation spaces $A_q^\gamma(L_p, \mathcal{T})$ is immediate from the inequalities (4.7) and (4.8), using the observation (2.31):

Theorem 4.3. *If $0 < \gamma < \alpha$ and $0 < q \leq \infty$, then*

$$A_q^\gamma(L_p, \mathcal{T}) = (L_p, B_\tau^\alpha(\mathcal{T}))_{\alpha, q}^\gamma$$

with equivalent norms.

The next result establishes an important (continuous) embedding, which will be needed in §5 in order to identify the approximation spaces (the ones determined by the algorithms, as well as best n -term Courant element approximation) as B -spaces.

Theorem 4.4. *Suppose our standing assumptions hold, i.e., $\alpha > 1$ if $p = \infty$ and $\alpha > 0$ if $p < \infty$. If we let $1/\tau := \alpha + 1/p$, then $A_\tau^\alpha(L_p, \mathcal{T}) \subset B_\tau^\alpha(\mathcal{T})$ and*

$$(4.9) \quad \|f\|_{B_\tau^\alpha(\mathcal{T})} \leq c\|f\|_{A_\tau^\alpha(L_p, \mathcal{T})},$$

where c depends only on α , p , and the parameters of \mathcal{T} .

Proof. We shall prove (4.9) only in the case $p = \infty$, proceeding similarly as in [7]. For a proof in the case $0 < p < \infty$, see [3]. Suppose $f \in A_\tau^\alpha(L_\infty, \mathcal{T})$, and let $S_m \in \Sigma_m(\mathcal{T})$ be such that

$$(4.10) \quad \|f - S_m\|_\infty \leq 2\sigma_m(f, \mathcal{T})_\infty.$$

Since $\sigma_m(f, \mathcal{T})_\infty \rightarrow 0$, we have $f = S_1 + \sum_{\nu=1}^{\infty} (S_{2^\nu} - S_{2^{\nu-1}})$ with the series converging uniformly, and hence ($\tau < 1$)

$$(4.11) \quad \|f\|_{B_\tau^\alpha(\mathcal{T})}^\tau \leq \|S_1\|_{B_\tau^\alpha(\mathcal{T})}^\tau + \sum_{\nu=1}^{\infty} \|S_{2^\nu} - S_{2^{\nu-1}}\|_{B_\tau^\alpha(\mathcal{T})}^\tau.$$

We apply the Bernstein estimate from Theorem 4.2 to $S_{2^\nu} - S_{2^{\nu-1}} \in \Sigma_{2^{\nu+1}}(\mathcal{T})$ to obtain

$$\|S_{2^\nu} - S_{2^{\nu-1}}\|_{B_\tau^\alpha(\mathcal{T})} \leq c2^{\nu\alpha} \|S_{2^\nu} - S_{2^{\nu-1}}\|_\infty \leq c2^{\nu\alpha} (\sigma_{2^\nu}(f, \mathcal{T})_\infty + \sigma_{2^{\nu-1}}(f, \mathcal{T})_\infty)$$

and similarly

$$\|S_1\|_{B_\tau^\alpha(T)} \leq c(\|f\|_\infty + \sigma_1(f, T)_\infty).$$

Substituting the above in (4.11), we find that

$$\|f\|_{B_\tau^\alpha(T)}^\tau \leq c\|f\|_\infty^\tau + c \sum_{\nu=1}^\infty (2^{\nu\alpha} \sigma_{2^\nu}(f, T)_\infty)^\tau \leq c\|f\|_{A_\tau^\alpha(L_\infty, T)}^\tau. \quad \square$$

5. APPROXIMATION SPACES FOR ALGORITHMS

Our goal in this section is to show that the algorithms that we developed and explored in §3 achieve (in a certain sense) the rate of convergence of the best n -term Courant element approximation. We shall utilize the characterization of the approximation spaces

$$A_q^\gamma(L_p, T; \sigma) := A_q^\gamma(L_p, T)$$

from the previous section (see Theorems 4.3 and 4.4). We shall denote by $A_q^\gamma(L_p, T; \mathbb{A}^T)$, $A_q^\gamma(L_p, T; \mathbb{A}^{TC})$, and $A_q^\gamma(L_p, T; \mathbb{A}^P)$ the approximation spaces generated by the “threshold”, “trim and cut”, and “push the error” algorithms, respectively. Namely, $f \in A_q^\gamma(L_p, T; \mathbb{A})$, where \mathbb{A} is \mathbb{A}^T , \mathbb{A}^{TC} or \mathbb{A}^P , if $f \in L_p(E)$ and

$$\|f\|_{A_q^\gamma(L_p, T; \mathbb{A})} := \|f\|_p + \left(\sum_{n=1}^\infty (n^\gamma \mathbb{A}_n(f, T)_p)^q \frac{1}{n} \right)^{1/q} < \infty,$$

with the usual modification when $q = \infty$ (it is not quite a norm).

Theorem 5.1. *Let T be an LR-triangulation of a bounded polygonal domain $E \subset \mathbb{R}^2$.*

(a) *If $p = \infty$, $\alpha > 1$, and $\tau := 1/\alpha$, then*

$$(5.1) \quad A_\tau^\alpha(L_\infty, T; \mathbb{A}^P) = A_\tau^\alpha(L_\infty, T; \mathbb{A}^{TC}) = A_\tau^\alpha(L_\infty, T; \sigma) = B_\tau^\alpha(T)$$

with equivalent “norms”.

(b) *If $0 < p < \infty$, $\alpha > 0$, and $\tau := (\alpha + 1/p)^{-1}$, then*

$$(5.2) \quad A_\tau^\alpha(L_p, T; \mathbb{A}^{TC}) = A_\tau^\alpha(L_p, T; \mathbb{A}^T) = A_\tau^\alpha(L_p, T; \sigma) = B_\tau^\alpha(T)$$

with equivalent “norms”, where “trim and cut” is applied with parameter $\tau \leq \varrho < p$.

Proof. (a) Let $p = \infty$. We let $\mathbb{A}_n(f)_\infty$ denote $\mathbb{A}_n^P(f)_\infty$ or $\mathbb{A}_n^{TC}(f)_\infty$, and $A_\tau^\alpha(L_\infty; \mathbb{A})$ denote the approximation space generated by the corresponding algorithm. Suppose $\|f\|_{A_\tau^\alpha(L_\infty; \mathbb{A})} < \infty$. Evidently, $\sigma_n(f)_\infty \leq \mathbb{A}_n(f)_\infty$, and hence, using Theorem 4.4,

$$\|f\|_{B_\tau^\alpha} \leq c\|f\|_{A_\tau^\alpha(L_\infty; \sigma)} \leq c\|f\|_{A_\tau^\alpha(L_\infty; \mathbb{A})}.$$

It remains to show that if $\|f\|_{B_\tau^\alpha} < \infty$, then

$$(5.3) \quad \|f\|_{A_\tau^\alpha(L_\infty; \mathbb{A})} \leq c\|f\|_{B_\tau^\alpha}.$$

For the proof of this estimate, we shall employ Lemmas 3.8 and 3.13. Since they are identical, it does not matter if we prove (5.3) for “push the error” or for “trim and cut”.

Suppose $f = \sum_{\theta \in \Theta} b_\theta \varphi_\theta$ is the representation of f that is used while “push the error” or “trim and cut” is applied. We have

$$\|f\|_{B_\tau^\alpha} := \left(\sum_{\theta \in \Theta} |b_\theta|^\tau \right)^{1/\tau}, \quad \tau := 1/\alpha, \quad \alpha > 1.$$

Next, we use a well-known interpolation technique. We choose $\alpha_0, \alpha_1, \tau_0$, and τ_1 as follows: $1 = \alpha_1 < \alpha < \alpha_0$ and $\tau_0 := 1/\alpha_0, \tau_1 := 1/\alpha_1$. Hence $0 < \tau_0 < \tau < \tau_1 = 1$. Now let $(|b_{\theta_j}|)_{j=1}^\infty$ be the decreasing rearrangement of the sequence $(|b_\theta|)_{\theta \in \Theta}$, i.e., indexed so that

(5.4)
$$|b_{\theta_1}| \geq |b_{\theta_2}| \geq \cdots.$$

We fix $\nu \geq 0$, and denote $f^0 := \sum_{j=1}^{2^\nu} b_{\theta_j} \varphi_{\theta_j}$ and $f^1 := \sum_{j=2^\nu+1}^\infty b_{\theta_j} \varphi_{\theta_j}$. In going further, we apply Lemma 3.8 or Lemma 3.13 to $f = f^0 + f^1$, from above, to obtain

$$\mathbb{A}_{2^\nu}(f)_\infty \leq c\, 2^{-\nu\alpha_0} \left(\sum_{j=1}^{2^\nu} |b_{\theta_j}|^{\tau_0} \right)^{1/\tau_0} + c\, 2^{-\nu} \sum_{j=2^\nu+1}^\infty |b_{\theta_j}|.$$

Using property (5.4) and the facts that $\tau = 1/\alpha, 1 < \alpha < \alpha_0$, and $\tau_0 = 1/\alpha_0$, we infer

$$\begin{aligned} \sum_{\nu=0}^\infty (2^{\nu\alpha} \mathbb{A}_{2^\nu}(f)_\infty)^\tau &\leq c \sum_{\nu=0}^\infty \left[2^{-\nu(\alpha_0-\alpha)\tau_0} \sum_{k=0}^\nu 2^k |b_{\theta_{2^k}}|^{\tau_0} \right]^{\tau/\tau_0} \\ &\quad + c \sum_{\nu=0}^\infty \left[2^{\nu(\alpha-1)} \sum_{k=\nu}^\infty 2^k |b_{\theta_{2^k}}| \right]^\tau \\ &\leq c \sum_{k=0}^\infty 2^k |b_{\theta_{2^k}}|^\tau \leq c \sum_{\theta \in \Theta} |b_\theta|^\tau = c \|f\|_{B_\tau^\alpha}^\tau, \end{aligned}$$

where we used the well-known Hardy inequalities, namely, we applied the inequality from Lemma 3.10 in [15] to estimate the first sum and Lemma 3.4 from [6] to the second term.

(b) For $0 < p < \infty$, the proof of (5.2) is similar to the proof of (5.1). The only difference is that the appropriate roles of Lemmas 3.8 or 3.13 are now played by Lemmas 3.2 or 3.10. We omit the details. □

6. CONCLUDING REMARKS

Our primary goal in the present article is to quantify the nonlinear n -term approximation from Courant elements and use it to develop algorithms capable of achieving the rate of the best approximation. This is closely related to the fundamental question in nonlinear approximation of how to measure the smoothness of the functions. As we show in this article, for n -term Courant element approximation when the triangulation \mathcal{T} is fixed, it is natural to measure the smoothness via the scale of the B -spaces $B_\tau^\alpha(\mathcal{T})$. The use of these spaces allows one to characterize the approximation spaces for any rate of convergence $\alpha > 0$. It also enables us to develop algorithms which attain the rate of the best approximation.

It is natural to add another degree of nonlinearity to the approximation by allowing the triangulation \mathcal{T} to vary. Thus a function f should be considered smooth of order $\alpha > 0$ if $\inf_{\mathcal{T}} \|f\|_{B_\tau^\alpha(\mathcal{T})} < \infty$, where the infimum is taken over all LR-triangulations \mathcal{T} (with fixed parameters). Therefore the rate of n -term Courant element approximation to f is roughly $O(n^{-\alpha})$. Summarizing, our approximation scheme proceeds as follows: (i) for a given function f , find a triangulation \mathcal{T}_f and

a B -space $B_\tau^\alpha(\mathcal{T}_f)$ in which f exhibits the most smoothness, (ii) find an optimal representation of f in terms of Courant elements from $\Phi_\mathcal{T}$, and (iii) run an algorithm that achieves the rate of the best n -term Courant element approximation. The first step in this scheme is the most complicated one. We do not have an efficient solution for this as yet. In the simpler case of nonlinear approximation by piecewise polynomials over dyadic partitions, this problem, however, has a complete and efficient solution [14]. As we show, once the triangulation \mathcal{T} is determined, the remaining two steps are now well understood and have efficient solutions in both theoretical and practical senses.

The three algorithms that we develop and explore in this article provide solutions of the problem under appropriate conditions. A common feature of these algorithms is the first step, a nontrivial decomposition from the redundant collection of all Courant elements from $\Phi_\mathcal{T}$. After this initial step, however, they take three different routes. The “threshold” algorithm is completely unstructured but easy to implement. The drawback of this procedure is that it is not valid in the case of the uniform norm, and as a consequence it does not perform well in L_p for p large. The “trim and cut” algorithm is valid for L_p , $0 < p \leq \infty$, but it is over-structured and as a result the performance suffers. The “push the error” algorithm appears to be the preferred approximation method.

The algorithms that we develop in this article are not restricted to n -term Courant element approximation. They can be applied immediately to the approximation from (discontinuous) piecewise approximation over multilevel triangulations (for the precise setting, see [11]). In this case the role of the B -spaces $B_\tau^\alpha(\mathcal{T})$ should be played by the skinny B -spaces $\mathcal{B}_\tau^\alpha(\mathcal{T})$, introduced in (2.37). The results are similar, but simplify considerably. We omit the details.

Furthermore, these algorithms can easily be adapted to nonlinear n -term approximation by smooth piecewise polynomial basis functions such as those considered in [3] and, in particular, by box splines. The main difference would be that one should use the corresponding B -spaces, developed in [3], but proceeding in a similar manner to this paper.

It is natural to use (wavelet or prewavelet) bases in nonlinear approximation, and specifically for approximation in L_p ($1 < p < \infty$). We are not aware of compactly supported wavelets (prewavelets) generated by Courant elements or smoother piecewise polynomials on general multilevel triangulations. It is clear to us that such wavelet bases would be very “expensive” to construct and hence are of limited practical value. However, in the case of uniform triangulations, compactly supported prewavelets and wavelet frames generated by Courant elements, or box splines, do exist and have been implemented in practice. Obviously, the n -term approximation from such bases or frames cannot surpass the rate of the best n -term Courant (or box spline) approximation, but they may give better constants and hence better performance results in practical situations.

It is also an important observation that, even in the case of uniform triangulations, the B -spaces used here are different from the Besov spaces used in nonlinear approximation. For a more complete discussion of this issue, see [11] and [3].

Finally, we remark that in a related paper [12] we extend the arguments of this paper to develop a corresponding approach in the Hausdorff metric which is natural for approximating surfaces. There we also consider various practical aspects for decompositions, numerical approximation, and data structures.

APPENDIX. COLORING LEMMA

In order to keep focus on the main analytical results of the paper, we have postponed the proof of the coloring lemma used in Section 3.2 to this appendix. This decomposition result was used to create a manageable collection of tree structures for estimating both the error and the number of elements used in our constructed approximant. Since this is a general purpose result which may prove useful in similar settings, we give its proof in full in this appendix. For clarity we have broken down the proof into a series of lemmas. Since the coloring is done in several refinement stages, it is helpful to think of the coloring as an ordered triple of *primary*, *secondary*, and *shade* colors. The primary coloring will sort the elements periodically by resolution level, the secondary coloring will insure there is spatial color separation, and the third coloring (shading) is a more delicate adjustment to insure that tree structures are formed. We begin by repeating the statement of the coloring lemma for the reader's convenience.

Coloring Lemma [see Lemma 3.2]. *For any LR-triangulation \mathcal{T} of E , the set $\Theta := \Theta(\mathcal{T})$ of all cells generated by \mathcal{T} can be represented as a finite disjoint union of its subsets $(\Theta^\nu)_{\nu=1}^K$ with $K = K(N_0, M_0)$ (N_0 is the maximal valence and M_0 is the maximal number of children of a triangle in \mathcal{T}) such that each Θ^ν has a tree structure with respect to the inclusion relation, i.e., if $\theta', \theta'' \in \Theta^\nu$, then $(\theta')^\circ \cap (\theta'')^\circ \neq \emptyset$, or $\theta' \subset \theta''$, or $\theta'' \subset \theta'$.*

To begin the proof, we show, without loss of generality, that for the purposes of coloring we may assume that the multiresolution triangulation provides sufficient resolution with each refinement step. We argue below that after a certain fixed number of increments of the level there will be a guaranteed refinement of each edge and triangle, which by hypothesis is controlled from above, i.e., uniformly bounded valences and max number of subtriangles for each refinement. Consequently, we may separate the levels of Θ into L ($L := \lceil 12N_0^4 \ln_2 M_0 \rceil$) disjoint classes (primary colors) by placing two levels in the same class iff their indices are the same (mod L). Thus a class $\tilde{\Theta}$ is of the form $\tilde{\Theta} = \bigcup_{j=0}^\infty \tilde{\Theta}_j$, where $\tilde{\Theta}_0 := \Theta_{j_0}$ for some $0 \leq j_0 < L$ and $\tilde{\Theta}_j := \Theta_{j_0+jL}$. Since each such class $\tilde{\Theta}$ has a different primary color, it will suffice to show how to designate the secondary colors of the members of a single $\tilde{\Theta}$. Therefore, to simplify the notation and wording of arguments, we will simply refer to (secondary) coloring the classes $\tilde{\Theta}$ instead of Θ . In Lemma A.1 below we show however that these classes have additional useful properties. Loosely speaking, part (a) shows that the old vertices on a given level are far apart in terms of the graph metric. In part (b) a similar statement is given for the “central parts” of non-overlapping edges of Courant elements from different levels of $\tilde{\Theta}$.

For $D \subset \mathbb{R}^2$ and $m \geq 0$, we define the *star* $St_m^k(D)$ inductively by $St_m^0(D) := D$ and $St_m^k(D) := \bigcup \{ \theta \in \Theta : \text{level } \theta = m, \theta^\circ \cap St_m^{k-1}(D) \neq \emptyset \}$. For the vertices in resolution level m , this is just the neighborhood of radius k in the graph metric. For an edge e with vertices v' and v'' and an integer $m > \text{level } e$, we define the “central part” of the edge to be $st(e, m) := St_m^2(e \setminus St_m^{R-1}(\{v', v''\}))$, where $R := M_0^{4N_0^2} + 4$. This selection for R has been made sufficiently large so that part (b) of the following lemma holds.

Lemma A.1. *The Courant collection $\tilde{\Theta}$ described above satisfies the following conditions:*

- (a) *For each edge $[v, v']$ the distance between v and v' , measured in the graph metric on the next finer level of $\tilde{\Theta}$, is at least $4R$.*
- (b) *If e and e' are edges from cells in $\tilde{\Theta}$, m is an integer with $m - L \geq \text{level } e \geq \text{level } e'$, and $e \not\subseteq e'$, then $st(e, m) \cap e' = \emptyset$.*

Proof. (a) Note that each edge in Θ gets subdivided at least once after $2N_0$ levels. Further, observe that after $\tilde{N}_0 := 2N_0^2$ refinements of any triangle, none of its vertices can be connected to their opposite edge by a single edge at the finer level. Using this observation repeatedly, one can verify that after L refinements, the graph metric distance between v and v' will be at least $2^{L/\tilde{N}_0} = M_0^{6N_0^2} > 4R$.

(b) Let v and v' be the vertices of e . Using twice the observation from the proof of part (a), we conclude that the distance from each of the vertices in $e \setminus St_{m-2\tilde{N}_0}^1(\{v, v'\})$ to e' is at least 4 when measured in the graph metric on level m . Therefore, on the m -th level, $e \setminus St_{m-2\tilde{N}_0}^1(\{v, v'\})$ has a buffer of at least three layers of triangles that separates it from e' . On the other hand, the existence of M_0 and the choice of R guarantee that $St_m^{R-1}(\{v, v'\}) \supset St_{m-2\tilde{N}_0}^1(\{v, v'\})$, and this establishes the claim. \square

This completes the primary coloring, and from this point on we only need work with a particular $\tilde{\Theta}$ (i.e., a fixed primary color). In this case “level θ ” will now refer to the level of θ in $\tilde{\Theta}$ rather than in Θ , as will the star $St_m^k(\theta)$ and $st(e, m)$. Also, when referring to the color of a cell we will now mean the secondary color, unless otherwise specified. For $\theta \in \tilde{\Theta}$ we denote by $\partial\theta$ the boundary of θ , and by x_θ the central point of θ . We say that the cells in $\Theta' \subset \tilde{\Theta}_m$ are R -disjoint ($R \geq 1$) if $\theta^\circ \cap St_m^R(\theta') = \emptyset$ for any $\theta, \theta' \in \Theta'$.

The next result is used for the (secondary) coloring of cells of $\tilde{\Theta}$, proceeding from coarse to fine levels, and uses M colors, so that same color cells are R -disjoint.

Lemma A.2. *Suppose some of the cells on a given level are colored in $M := N_0^{R+1} + 1$ ($R \geq 1$) colors so that the same color cells are R -disjoint. Then the rest of the cells on that level can be colored in the same M colors so that the same color cells are R -disjoint.*

Proof. To complete the coloring on the given level, we first use color #1 to paint as many cells as possible so that the same color cells are R -disjoint. Next, we use color #2 as much as possible, followed by the third and so on until either all cells get painted or we run out of colors. The latter case, however, never occurs. Indeed, assume the contrary and let θ be the first cell that cannot be colored by this algorithm with the M colors. The cell θ has the property that within its $R+1$ star $St_m^{R+1}(\theta)$ there must be at least one cell painted with each of the M colors. But this contradicts the fact that M was selected to be at least as large as the number of cells within $St_m^{R+1}(\theta)$. \square

For the secondary coloring we proceed inductively, beginning at the coarsest level $\tilde{\Theta}_0$, and color cells in M colors so that same color cells are R -disjoint. Suppose then that all levels up to $\tilde{\Theta}_k$ ($k > 0$) have been colored. We color $\tilde{\Theta}_k$ as follows.

Step a) (Color corner cells). First we define the notion of corner cell. A cell θ of level k is called a *corner cell* for a coarser cell θ' if θ' has an adjacent cell θ'' (at

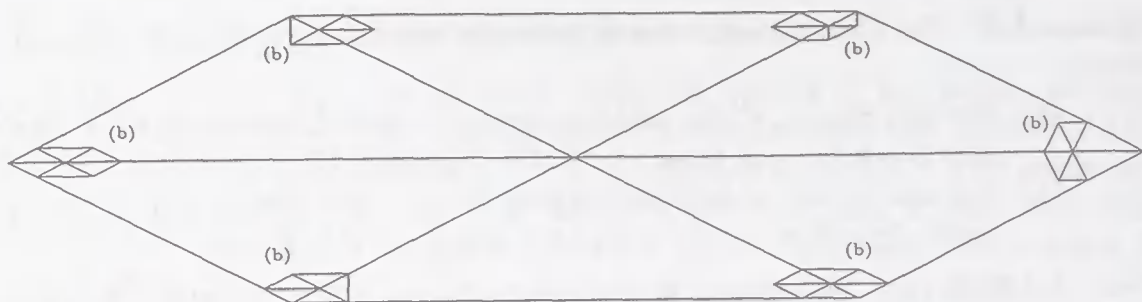


FIGURE 2. Corner cells from Step a)

the same level of course) so that x_θ lies on edge $[x_{\theta'}, x_{\theta''}]$ and x_θ is adjacent to $x_{\theta''}$ on the level k (see Figure 2). Given a cell $\theta' \in \tilde{\Theta}_{k-1}$, we color each of its corner cells $\theta \in \tilde{\Theta}_k$ the same color as θ' . This insures that a cell's color is propagated through all finer levels to its corner cells.

Step b) (*Extend the coloring to R -stars of the vertices on level $(k-1)$*). For each vertex v on level $(k-1)$, we paint the cells contained in $St_k^{R+2}(v)$ using M colors so that the coloring done in Step a) is preserved and each color is used at most once. This is always possible, since M was selected sufficiently large. Note that after this step the same color cells are R -disjoint, since part (a) of Lemma A.1 guarantees that the stars are sufficiently separated.

Step c) (*Complete the secondary coloring of $\tilde{\Theta}_k$*). Accounting for the cells previously painted in Steps a) and b), we color the remaining cells from $\tilde{\Theta}_k$ as described in Lemma A.2.

This procedure specifies the secondary coloring of $\tilde{\Theta}$, and we have thus represented it as a finite disjoint union $\bigcup_{\nu=1}^M \tilde{\Theta}^\nu$, where $\tilde{\Theta}^\nu$ are all cells (secondarily) colored in the ν -th color. Thus the primary color skips levels until sufficient refinement is guaranteed, while the secondary color insures sufficient spatial separation on each level to control cell overlaps. Unfortunately, the collection of same primary-secondary colored cells ($\tilde{\Theta}^\nu$) might not form a tree structure, i.e., there might be two cells in $\tilde{\Theta}^\nu$ whose interiors meet but neither of them contains the other. This may only happen when a finer cell lies on the edge of a given cell. To fix this defect we will set for each fixed $\tilde{\Theta}^\nu$ the third coloring component, the *shade* of the cells, from two possible choices. First, we say that θ' and θ'' ($\theta', \theta'' \in \Theta$) *touch* if an edge of the finer of the cells is contained in an edge of the coarser.

We now restrict our cells to be of fixed primary and secondary colors (i.e., fix $\tilde{\Theta}^\nu$) and inductively determine the *shade* of these cells. On the coarsest level $\tilde{\Theta}_0^\nu$ of $\tilde{\Theta}^\nu$ all cells are disjoint, and we assign them shade $\#1$. For the induction step, we suppose cells of all levels of $\tilde{\Theta}^\nu$ up to level k have been shaded and each shaded collection satisfies the desired tree properties. We say that a cell θ is *shade-consistent* with $\tilde{\theta}$ if x_θ does not lie on an edge of any cell that has the same shade as $\tilde{\theta}$. Hence it is possible to place θ in this shade collection and preserve the tree structure. In this case we will also use the terminology that θ is consistent with that particular shade. We now proceed to shade the cells belonging to level k , i.e., $\theta \in \tilde{\Theta}_k^\nu$, according to:

Case i. If θ both touches and is shade-consistent with some coarser cell $\tilde{\theta}$, then we assign to θ the same shade as that of the finest such $\tilde{\theta}$. Recall that this finest cell is unique by the construction of $\tilde{\Theta}^\nu$.

Case ii. Otherwise, we assign to θ the first numbered shade for which θ is consistent. If no such shade exists, we introduce a new shade for θ .

By the construction in the induction step, it is obvious that each shade subcollection has the desired tree structure. We will show that these criteria introduce at most two shades. For this we need a couple of technical facts. We remind the reader that all cells belong to a fixed $\tilde{\Theta}^\nu$, i.e., they have a fixed primary and secondary color.

Lemma A.3. *If θ intersects an edge e' of a coarser cell θ' but is not one of its corner cells, then $\theta \subset st(e', m)$, where $m := \text{level } \theta$.*

Proof. Let e' be an edge of θ' that intersects θ , and let v be a vertex of e' . By Step b) of our coloring procedure (for secondary colors), $St_m^1(v)$ contains a corner cell θ'' in $\tilde{\Theta}_m^\nu$ that is shaded the same as θ' . By Step c) in the construction of $\tilde{\Theta}^\nu$, $St_m^R(v)$ does not contain any other cells from $\tilde{\Theta}_m^\nu$. Since θ is not a corner cell of θ' , then $\theta \neq \theta''$. Therefore $\theta \cap St_m^{R-1}(v) = \emptyset$, and so θ must meet $e' \setminus St_m^{R-1}(\{v, v'\})$, where v' is the remaining vertex of e' . Therefore, $\theta \subset st(e', m)$. \square

Lemma A.4. *Cells of $\tilde{\Theta}^\nu$ with different shades do not touch.*

Proof. Suppose to the contrary that cells $\theta_j, \theta_k \in \tilde{\Theta}^\nu$ of different shades (shade $\#j$, shade $\#k$, respectively) do touch. We may first assume that θ_j is a maximal (i.e., coarsest level) cell of shade $\#j$ that touches θ_k , and conversely, that θ_k is a maximal cell in shade $\#k$ that touches θ_j . This follows by iteration and the fact that there are only finitely many coarser levels; so the iteration must terminate.

We may assume without loss of generality that $\text{level } \theta_j < \text{level } \theta_k =: m_k$, and let e_j, e_k denote the edges of θ_j, θ_k respectively, such that $e_k \subset e_j$. We consider the two cases under which the finer cell θ_k could have been shaded, and show that each one leads to a contradiction.

For Case i. In this event there would be a coarser cell $\tilde{\theta}_k \in \tilde{\Theta}^\nu$ of shade $\#k$ that touches θ_k and to which θ_k would be shade-consistent. Let \tilde{e}_k be an edge of $\tilde{\theta}_k$ where it is touched by θ_k . We consider two possible subcases, depending upon the relative level of $\tilde{\theta}_k$ to that of θ_j .

Subcase i.a $\tilde{\theta}_k$ is finer than θ_j .

Since $\text{level } \theta_j \leq \text{level } \tilde{\theta}_k < \text{level } \theta_k$, then by part (b) of Lemma A.1 either $\tilde{e}_k \subset e_j$ or $st(\tilde{e}_k, m_k) \cap e_j = \emptyset$. The first possibility may be ruled out, since it would imply that the coarser cell $\tilde{\theta}_k$ would touch θ_j , but θ_k is the maximal such cell of shade $\#k$. Hence $st(\tilde{e}_k, m_k)$ must be disjoint from e_j . Note that θ_k is not a corner cell of $\tilde{\theta}_k$. If that were the case, then θ_k would be disjoint from the interiors of all edges on level $\tilde{\theta}_k$ except the edge on which x_{θ_k} lies and the edges (at most two, possibly one) where $\tilde{\theta}_k$ is touched by θ_k . Hence, e_j must overlie one of these edges, since it contains e_k . This, however, contradicts the fact that θ_j touches θ_k in the former case and contradicts the maximality of θ_k in the latter. Therefore θ_k cannot be a corner cell of $\tilde{\theta}_k$, and so, by Lemma A.3, $\theta_k \subset st(\tilde{e}_k, m_k)$. But we have already proved that $st(\tilde{e}_k, m_k) \cap e_j = \emptyset$, which is impossible, since θ_k touches θ_j on e_j .

Subcase i.b $\tilde{\theta}_k$ is coarser than θ_j .

Since $\text{level } \tilde{\theta}_k < \text{level } \theta_j < \text{level } \theta_k$, then again by part (b) of Lemma A.1 either $e_j \subset \tilde{e}_k$ or $st(e_j, m_k) \cap \tilde{e}_k = \emptyset$. The former case contradicts maximality of θ_k relative to θ_j . For the latter case, note that θ_k cannot be a corner cell of θ_j , because θ_k

and θ_j have different shades. Therefore, by Lemma A.3, $\theta_k \subset st(e_j, m_k)$, and so we obtain $\theta_k \cap \tilde{e}_k = \emptyset$, which is impossible, since θ_k touches $\tilde{\theta}_k$ on \tilde{e}_k .

For Case ii. If this case occurred for the shading of θ_k , then since θ_j is both coarser than and touches θ_k , θ_k must not have been shade $\#j$ consistent. Hence there must be a $\tilde{\theta}_j \in \tilde{\Theta}^\nu$ of shade $\#j$ that is coarser than θ_k , and x_{θ_k} belongs to some edge \tilde{e}_j of $\tilde{\theta}_j$. We consider two possible subcases, depending upon the level of $\tilde{\theta}_j$ relative to that of θ_j .

Subcase ii.a θ_j is coarser than $\tilde{\theta}_j$.

Since level $\theta_j \leq \text{level } \tilde{\theta}_j < \text{level } \theta_k$, then compare edges e_j, \tilde{e}_j using part (b) of Lemma A.1 to infer either $st(\tilde{e}_j, m_k) \cap e_j = \emptyset$ or $\tilde{e}_j \subset e_j$. In the latter case, it follows that both the edge e_k (recall θ_k touches the coarser θ_j on e_k) and the opposite vertex x_{θ_k} (since $x_{\theta_k} \in \tilde{e}_j$) of a triangle in \mathcal{T}_k are contained in e_j , which is clearly impossible. If the former case holds, i.e., $st(\tilde{e}_j, m_k) \cap e_j = \emptyset$, then a contradiction also results. To see this, observe that θ_k cannot be a corner cell for $\tilde{\theta}_j$, due to the fact that they have different shades. But Lemma A.3 implies that $\theta_k \subset st(\tilde{e}_j, m_k)$, which contradicts the fact that $\theta_k \cap e_j \neq \emptyset$.

Subcase ii.b θ_j is finer than $\tilde{\theta}_j$.

Since level $\tilde{\theta}_j < \text{level } \theta_j < \text{level } \theta_k$, we again compare edges \tilde{e}_j, e_j using part (b) of Lemma A.1 to imply either $st(e_j, m_k) \cap \tilde{e}_j = \emptyset$ or $e_j \subset \tilde{e}_j$. By quite similar arguments to the previous subcase we can prove that contradictions are reached. Specifically, the latter statement implies that both the central vertex x_{θ_k} and its opposite edge e_k belong to the edge \tilde{e}_j . On the other hand, the fact that θ_k cannot be a corner cell for θ_j will imply that $\theta_k \subset st(e_j, m_k)$, which will show that x_{θ_k} belongs to the intersection $st(e_j, m_k) \cap \tilde{e}_j$, and contradict the former statement above.

By our assumption that different shaded cells could touch, we are led in all cases to contradictions, thereby completing our contrapositive proof. \square

By combining the previous results with the next lemma, it follows immediately that Θ can be colored with $K := 2ML$ colors, and the proof of the coloring lemma will be complete.

Lemma A.5. *At most two shades are required.*

Proof. Suppose in Case ii of the shading step above that a third shade were needed for some cell θ . Then its central point $x_\theta \in e_1 \cap e_2$ for some edges e_1 of θ_1 and e_2 of θ_2 , where $\theta_1, \theta_2 \in \tilde{\Theta}^\nu$ are coarser than θ and have shade $\#1$ and shade $\#2$, respectively. Now, if x_θ were a vertex for e_1 , then there would be a corner cell of θ_1 in $\tilde{\Theta}^\nu$ adjacent to θ , which is clearly impossible, since cells at the same level are R -disjoint. The same reasoning applies to e_2 . Therefore x_θ cannot be a vertex for either e_1 or e_2 , and we conclude that $e_1^\circ \cap e_2^\circ \neq \emptyset$. Hence, θ_1 and θ_2 touch, which contradicts Lemma A.4. \square

REFERENCES

- [1] C. Bennett and R. Sharpley, *Interpolation of operators*, Pure and Applied Mathematics Vol. **129**, Academic Press, Inc., Boston, MA, 1988. MR **89e**:46001
- [2] J. Bergh and J. Löfström, *Interpolation spaces: An introduction*, Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976. MR **58**:2349
- [3] O. Davydov and P. Petrushev, Nonlinear approximation from differentiable piecewise polynomials, 2002, preprint.

- [4] R. DeVore, I. Daubechies, A. Cohen, and W. Dahmen, Tree approximation and optimal encoding, *Appl. Comput. Harmon. Anal.* II (2001), 192–226. MR **2002g**:42048
- [5] R.A. DeVore, B. Jawerth, and B. Lucier, Surface compression, *Computer Aided Geometric Design* **9** (1992), 219–239. MR **93i**:65029
- [6] R.A. DeVore and G.G. Lorentz, *Constructive Approximation*, Grundlehren der Mathematischen Wissenschaften, Vol. **303**, Springer-Verlag, Heidelberg, 1993. MR **95f**:41001
- [7] R.A. DeVore, P. Petrushev, and X. Yu, Nonlinear wavelet approximation in the space $C(R^d)$, *Progress in Approximation Theory* (A. A. Gonchar, E. B. Saff, eds.), Springer-Verlag, New York, 1992, pp. 261–283. MR **94h**:41070
- [8] R.A. DeVore and V. Popov, Interpolation of Besov spaces, *Trans. Amer. Math. Soc.* **305**(1988), 397–414. MR **89h**:46044
- [9] R.A. DeVore and V. Popov, Interpolation spaces and non-linear approximation, in *Function Spaces and Applications*, M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin (eds.), Springer Lecture Notes in Math. **1302**, Springer-Verlag, Berlin, 1988, 191–205. MR **89d**:41035
- [10] M.A. Duchaineau, M. Wolinsky, D.E. Sigeti, M.C. Miller, C. Aldrich, and M.B. Mineev-Weinstein, ROAMing Terrain: Real-time Optimally Adapting Meshes, *Proc. IEEE Visualization '97*, October 1997, pp. 81–88.
- [11] B. Karaivanov and P. Petrushev, Nonlinear piecewise polynomial approximation beyond Besov spaces, 2001, preprint. (<http://www.math.sc.edu/~imip/01.html>).
- [12] B. Karaivanov, P. Petrushev and R.C. Sharpley, Algorithms for nonlinear piecewise polynomial approximation, 2002, preprint.
- [13] P. Petrushev, Direct and converse theorems for spline and rational approximation and Besov spaces, in *Function Spaces and Applications*, M. Cwikel et. al. (eds), Vol. 1302 of Lecture Notes in Mathematics, Springer, Berlin, 1988, pp. 363–377. MR **89d**:41027
- [14] P. Petrushev, Multivariate n -term rational and piecewise polynomial approximation, *J. Approx. Theory* (2003), to appear. (<http://www.math.sc.edu/~imip/01.html>)
- [15] P. Petrushev and V. Popov, *Rational approximation of real functions*, Cambridge University Press, 1987. MR **89i**:41022

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA, SOUTH CAROLINA 29208

E-mail address: `karaivan@math.sc.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA, SOUTH CAROLINA 29208

E-mail address: `pencho@math.sc.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTH CAROLINA, COLUMBIA, SOUTH CAROLINA 29208

E-mail address: `sharples@math.sc.edu`

THE ALMOST-DISJOINTNESS NUMBER MAY HAVE COUNTABLE COFINALITY

JÖRG BRENDLE

ABSTRACT. We show that it is consistent for the almost-disjointness number \mathfrak{a} to have countable cofinality. For example, it may be equal to \aleph_ω .

INTRODUCTION

Cardinal invariants of the continuum, that is, cardinal numbers between \aleph_1 and \mathfrak{c} (the size of the continuum) which are defined as the smallest size of a family of real numbers with a certain combinatorial property, play an increasingly important role in modern set theory. Equalities and inequalities between cardinal invariants have many connections with problems arising naturally in general topology, real analysis and algebra, and, from a purely set-theoretic point of view, there is a deep interplay with forcing theory, in particular in the light of the search for new iteration techniques.

One of the most basic questions about cardinal invariants is which values they can assume, and, for almost all cardinals, it is known that any regular value is possible.¹ Furthermore, most cardinals can either be shown to be regular in *ZFC* or they are equal to \mathfrak{c} in the random real model, in the Cohen real model,² or even in both, so that they can be consistently singular of uncountable cofinality. Notable exceptions are the splitting number \mathfrak{s} of which it is still unknown whether it may be singular [V] and the almost-disjointness number \mathfrak{a} , which has recently been shown to be consistently singular of uncountable cofinality by Shelah [S2].³ Things get trickier when one considers singular cardinals of countable cofinality. In fact, by far most of the cardinals, even those singular in the Cohen or random models, can be shown to have uncountable cofinality in *ZFC*.⁴ Exceptions are

- Shelah [S1] has proved that the covering number of the null ideal may have countable cofinality,
- the almost-disjointness number \mathfrak{a} dealt with below,

Received by the editors October 3, 2001.

2000 *Mathematics Subject Classification*. Primary 03E17; Secondary 03E35.

Key words and phrases. Maximal almost-disjoint families, almost-disjointness number, iterated forcing.

Supported by Grant-in-Aid for Scientific Research (C)(2)12640124, Japan Society for the Promotion of Science.

¹In fact, most of the cardinals that have been studied are equal to \mathfrak{c} under Martin's axiom *MA*.

²The models obtained by adding $\kappa = \kappa^\omega$ many random or Cohen reals.

³Cardinals relevant for this paper will be defined below. Also see [vD], [V], or [Bl].

⁴This is also true for \mathfrak{s} .

- it is unknown whether the reaping number \mathfrak{r} [M, Problem 3.4] or the independence number \mathfrak{i} can have countable cofinality.⁵

Here we show

Main Theorem. *Assume CH and let λ be a singular cardinal of countable cofinality. Then there is a forcing extension satisfying $\mathfrak{a} = \lambda$. In particular, $\mathfrak{a} = \aleph_\omega$ is consistent.*

We give a brief outline of the proof. It is well known that, assuming CH , one can force a mad family of size \aleph_ω such that, by a standard isomorphism-of-names argument, there is no mad family of size \aleph_n for $n \geq 2$ in the generic extension [H1]. However, mad families of size \aleph_1 may survive the forcing. The simplest way to get rid of small mad families is by iteratively adding dominating reals, say for \aleph_2 steps. If this is done in such a way that every dominating real is dominating only over a fragment of the mad family of size \aleph_ω added in the initial step, the latter will survive. Yet, if the dominating reals are added in a standard way, the forcing will lose most of its homogeneity, the isomorphism-of-names argument will cease to work, and there may be a mad family of size \aleph_2 instead. This is where Shelah's recent technique of iteration along templates [S2] comes in.⁶ It provides for a way of adjoining dominating reals with a forcing having enough local homogeneity. However, since isomorphism-of-names arguments require CH in the ground model, we need to describe the two-step extension sketched above in one step and incorporate the forcing adding the mad family of size \aleph_ω into the template framework. This accounts for some of the technical difficulties described below.

Apart from proving the Main Theorem, we also present a new, more axiomatic, treatment of the template framework (Section 1). While Shelah [S2] defines the template via two procedures building more complicated sets (with larger depth) from simpler ones, we only require that the template is a family of subsets of the linear order underlying the iteration satisfying several axioms, most notably well-foundedness. We think this approach is more lucid, apart from providing for a simpler definition of the iteration. There is a price one has to pay for that, however: the Completeness Lemma 1.1, showing that we are indeed dealing with an iteration, requires some additional work. Of course, our general approach can also be used to prove Shelah's original results [S2] (see [Br]) and, since there is no need to add a mad family in this case, definitions will be simpler than in Section 1.

In Section 2 we describe the template used for the proof of the Main Theorem, and in Section 3 we provide the isomorphism-of-names argument needed to complete the proof.

The template framework developed in Section 1 for Hechler forcing can in fact be used to handle a large class of easily definable ccc forcing notions (see [Br, Section 4] for a few examples). Replacing the forcing generically adjoining a mad family by an appropriate relative, one can get analogous results for relatives of \mathfrak{a} , e.g., for \mathfrak{a}_s , the size of the smallest mad family of partial functions from ω to ω . So, for example, $\mathfrak{a}_s = \aleph_\omega$ is consistent, and so is $\mathfrak{a} = \aleph_1 < \mathfrak{a}_s = \aleph_\omega$. For the latter

⁵As for maximal almost-disjoint families, it is rather easy to force a maximal independent family of size, say, \aleph_ω . Both \mathfrak{r} and \mathfrak{i} are equal to \mathfrak{c} , and thus possibly singular, in the Cohen and random models.

⁶As mentioned above, Shelah showed among other things that \mathfrak{a} could be singular. In his models, \mathfrak{a} is equal to \mathfrak{c} and therefore has uncountable cofinality.

result, one needs to replace Hechler forcing by eventually-different-reals forcing in the framework of Section 1 (cf. [Br, Section 4]).

Let us briefly recall the main notions relevant for this paper. Two infinite subsets A and B of ω are called *almost-disjoint* if their intersection is finite. $\mathcal{A} \subseteq [\omega]^\omega$ is an *almost-disjoint family* if its members are pairwise almost disjoint. \mathcal{A} is a *mad family* (*maximal almost-disjoint family*) if it is maximal with respect to being an almost-disjoint family, i.e., for every $B \in [\omega]^\omega$ there is $A \in \mathcal{A}$ such that $A \cap B$ is infinite. The *almost-disjointness number* \mathfrak{a} is the size of the least infinite mad family. For functions $f, g \in \omega^\omega$, we say that g *eventually dominates* f (and write $f \leq^* g$) if the set $\{n; f(n) > g(n)\}$ is finite. The *unbounding number* \mathfrak{b} is the cardinality of the smallest unbounded family in the structure (ω^ω, \leq^*) , that is, the size of the smallest $\mathcal{F} \subseteq \omega^\omega$ such that for all $g \in \omega^\omega$ there is $f \in \mathcal{F}$ with $f \not\leq^* g$. The *dominating number* \mathfrak{d} is the size of the least cofinal family in (ω^ω, \leq^*) . It is well known and easy to see that $\mathfrak{b} \leq \mathfrak{d}$ and $\mathfrak{b} \leq \mathfrak{a}$ in ZFC [vD].

Hechler forcing [H2] \mathbb{D} (see also [BJ]) consists of pairs (s, f) where $s \in \omega^{<\omega}$, $f \in \omega^\omega$ and $s \subseteq f$, ordered by $(t, g) \leq (s, f)$ if $t \supseteq s$, and $g \geq f$ everywhere. It generically adds a *dominating real*, that is, a real that eventually dominates all ground model reals. It is this forcing adjoining a dominating real which we shall use in the template framework sketched above.

Our notation is standard. For cardinal invariants of the continuum, we refer to [vD], [V] or [Bl]. For forcing theory, in particular for forcing related to cardinal invariants of the continuum, see [BJ].

I thank Juris Steprāns for pointing out a flaw in an earlier version of this work.

1. TEMPLATES AND ITERATIONS

The most useful definition of a template seems to be (see also [Br]) the following.

Definition (Template). A *template* is a pair (L, \mathcal{I}) such that (L, \leq) is a linear order and $\mathcal{I} \subseteq \mathcal{P}(L)$ is a family of subsets of L satisfying

- (1) $\emptyset, L \in \mathcal{I}$,
- (2) \mathcal{I} is closed under finite unions and intersections,
- (3) if $y < x$ belong to L , then there is $A \in \mathcal{I} \cap \mathcal{P}(L_x)$ such that $y \in A$,
- (4) if $A \in \mathcal{I}$ and $x \in L \setminus A$, then $A \cap L_x \in \mathcal{I}$,
- (5) \mathcal{I} is well-founded, i.e., there is a function $\text{Dp} = \text{Dp}_{\mathcal{I}} : \mathcal{I} \rightarrow \text{On}$, called *depth*, recursively defined by $\text{Dp}(\emptyset) = 0$ and $\text{Dp}(A) = \sup\{\text{Dp}(B) + 1; B \in \mathcal{I} \text{ and } B \subset A\}$ for $A \in \mathcal{I} \setminus \{\emptyset\}$.

Here, $L_x = \{y \in L; y < x\}$ is the *restriction* of L to x . If $A \subseteq L$, we define $\mathcal{I} \restriction A = \{B \cap A; B \in \mathcal{I}\}$, the *trace* of \mathcal{I} on A . $(A, \mathcal{I} \restriction A)$ is also a template.

Since \mathcal{I} is closed under finite intersections, if $A \in \mathcal{I}$, then $\mathcal{I} \restriction A = \{B \in \mathcal{I}; B \subseteq A\}$.

In our context, we need to slightly revise this definition because we will not only have “iteration coordinates” L_{Hech} used for adjoining Hechler generics but also “product coordinates” L_{mad} used for adding a mad family. Since the former should be generic over some, but not all, of the latter, we need to incorporate them into the same template framework L , and some of the above clauses should be true for all of L , while others need to be satisfied only for members of L_{Hech} .

Accordingly, let L_{Hech} and L_{mad} be disjoint sets, put $L = L_{\text{Hech}} \cup L_{\text{mad}}$, and assume L is equipped with a linear order. Further suppose $\mathcal{I} \subseteq \mathcal{P}(L)$ satisfies, in addition to (1) and (2) above, the following clauses:

- (3') if $x \in L_{\text{Hech}}$ and $y \in L_x$, then there is $A \in \mathcal{I} \cap \mathcal{P}(L_x)$ such that $y \in A$;
- (4') if $A \in \mathcal{I}$ and $x \in L_{\text{Hech}} \setminus A$, then $A \cap L_x \in \mathcal{I}$;
- (5') the trace $\mathcal{I} \upharpoonright L_{\text{Hech}} = \{A \cap L_{\text{Hech}}; A \in \mathcal{I}\}$ is well-founded,

as well as

- (6) if $A \in \mathcal{I}$ and $x \in A$, then $L_x \cap L_{\text{mad}} \subseteq A$.

This is the definition of “template” we shall work with for the remainder of the paper. Notice that (5') means in particular that the depth function Dp depends only on the L_{Hech} -part, i.e., $\text{Dp}(A) = 0$ iff $A \subseteq L_{\text{mad}}$ and, recursively, $\text{Dp}(A) = \sup\{\text{Dp}(B) + 1; B \in \mathcal{I} \text{ and } B \cap L_{\text{Hech}} \subset A \cap L_{\text{Hech}}\}$.

(6) is a closure condition for the L_{mad} -part which is needed to make the proof of Main Lemma 1.1 below go through. More generally, we say that $A \subseteq L$ is *closed* if A satisfies (6). (So \mathcal{I} consists only of closed sets.) For arbitrary $A \subseteq L$, we then define its *closure* $\text{cl}(A)$ by $\text{cl}(A) = A \cup \bigcup_{x \in A} (L_x \cap L_{\text{mad}})$. Thus $\text{cl}(A)$ is the smallest closed set containing A and $\text{cl}(\text{cl}(A)) = \text{cl}(A)$.

The basic idea for the following, attempted, definition comes from [S2]. It is modified, however, due to our axiomatic treatment of the concept of “template” (see also [Br]) and because of the inclusion of L_{mad} .

Definition (Iterating Hechler forcing and adding a mad family along a template). Assume (L, \mathcal{I}) is as above. We define, for $A \in \mathcal{I}$, by recursion on $\text{Dp}(A)$, the partial order (p.o.) $\mathbb{P} \upharpoonright A$ (more explicitly, we define $\mathbb{P} \upharpoonright (A, \mathcal{I})$, but we shall drop the reference to \mathcal{I} in case there is no ambiguity).

- $\text{Dp}(A) = 0$. This means $A \subseteq L_{\text{mad}}$. $\mathbb{P} \upharpoonright A$ consists of all finite partial functions p with domain contained in A and such that $p(z) \in 2^n$ for all $z \in \text{dom}(p)$ for some $n = n^p \in \omega$. The ordering on $\mathbb{P} \upharpoonright A$ is given by: $q \leq_{\mathbb{P} \upharpoonright A} p$ if $\text{dom}(q) \supseteq \text{dom}(p)$ and
 - $n^q \geq n^p$, $p(z) \subseteq q(z)$ for all $z \in \text{dom}(p)$, $|\{z \in \text{dom}(p); q(z)(i) = 1\}| \leq 1$ for all $i \in n^q \setminus n^p$.
- $\text{Dp}(A) > 0$. $\mathbb{P} \upharpoonright A$ consists of all finite partial functions p with domain contained in A and such that
 - there is $n = n^p \in \omega$ with $p(z) \in 2^n$ for all $z \in \text{dom}(p) \cap L_{\text{mad}}$;
 - letting $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$, there is $B \in \mathcal{I} \cap \mathcal{P}(A \cap L_x)$ (so $\text{Dp}(B) < \text{Dp}(A)$) such that $p \upharpoonright (A \cap L_x) \in \mathbb{P} \upharpoonright B$ and $p(x) = (s_x^p, f_x^p)$ where $s_x^p \in \omega^{<\omega}$ and f_x^p is a $\mathbb{P} \upharpoonright B$ -name for an element of ω^ω such that $p \upharpoonright (A \cap L_x) \Vdash_{\mathbb{P} \upharpoonright B} “s_x^p \subseteq \dot{f}_x^p”$ (this means $p(x)$ is a $\mathbb{P} \upharpoonright B$ -name⁷ for a condition in Hechler forcing \mathbb{D}).

The ordering on $\mathbb{P} \upharpoonright A$ is given by: $q \leq_{\mathbb{P} \upharpoonright A} p$ if $\text{dom}(q) \supseteq \text{dom}(p)$ and

- $n^q \geq n^p$, $p(z) \subseteq q(z)$ for all $z \in \text{dom}(p) \cap L_{\text{mad}}$, $|\{z \in \text{dom}(p) \cap L_{\text{mad}}; q(z)(i) = 1\}| \leq 1$ for all $i \in n^q \setminus n^p$ (this guarantees that the reals added in coordinates from L_{mad} are characteristic functions of an almost-disjoint family), as well as

⁷Note, however, that the first coordinate s_x^p of the condition $p(x)$ is not a name.

- either $y = \max(\text{dom}(q) \cap L_{\text{Hech}}) > x = \max(\text{dom}(p) \cap L_{\text{Hech}})$ and there is $B \in \mathcal{I} \cap \mathcal{P}(A \cap L_y)$ such that $p \restriction (A \cap L_y), q \restriction (A \cap L_y) \in \mathbb{P} \restriction B$ and $q \restriction (A \cap L_y) \leq_{\mathbb{P} \restriction B} p \restriction (A \cap L_y)$,
- or $x = \max(\text{dom}(q) \cap L_{\text{Hech}}) = \max(\text{dom}(p) \cap L_{\text{Hech}})$ and there is $B \in \mathcal{I} \cap \mathcal{P}(A \cap L_x)$ such that $p \restriction (A \cap L_x), q \restriction (A \cap L_x) \in \mathbb{P} \restriction B$, $q \restriction (A \cap L_x) \leq_{\mathbb{P} \restriction B} p \restriction (A \cap L_x)$, \dot{f}_x^p, \dot{f}_x^q are $\mathbb{P} \restriction B$ -names, $s_x^p \subseteq s_x^q$, and $q \restriction (A \cap L_x) \Vdash_{\mathbb{P} \restriction B} \text{“}\dot{f}_x^p(n) \leq \dot{f}_x^q(n) \text{ for all } n\text{”}$ (the last two clauses mean that $q \restriction (A \cap L_x) \Vdash_{\mathbb{P} \restriction B} \text{“}q(x) \leq_{\mathbb{D}} p(x)\text{”}$).

We have not argued yet that this recursive definition works at all. The point is this requires that all $\mathbb{P} \restriction A$'s be transitive, which is not trivial because the sets $B \in \mathcal{I}$ witnessing that $q \leq p$ may depend on the pair (p, q) . Therefore, to prove transitivity, we need to show that the $\mathbb{P} \restriction B$ completely embed one into the other. This will be done in Main Lemma 1.1 below.

Note that, once this is achieved, $\mathbb{P} \restriction (A, \mathcal{I}) = \mathbb{P} \restriction (A, \mathcal{I} \restriction A)$ is immediate for $A \in \mathcal{I}$. Of course, the above recursion also defines $\mathbb{P} \restriction A = \mathbb{P} \restriction (A, \mathcal{I} \restriction A)$ for arbitrary $A \subseteq L$.

If $A, B \in \mathcal{I}$, $A \subset B$, then $\mathbb{P} \restriction A \subset \mathbb{P} \restriction B$ is immediate from the definition (because $\mathcal{I} \restriction A \subset \mathcal{I} \restriction B$ in this case). This is much less clear if one of A or B does not belong to \mathcal{I} . Neither is it clear whether $\mathbb{P} \restriction A <_{\circ} \mathbb{P} \restriction B$, the most basic property the above recursive definition must satisfy to make it an iteration, even in case both A and B come from \mathcal{I} . This issue is addressed by the following crucial lemma.

Main Lemma 1.1 (Completeness of embeddings). *Let $B \in \mathcal{I}$ and $A \subset B$ be closed. Then $\mathbb{P} \restriction B$ is a partial order, $\mathbb{P} \restriction A \subset \mathbb{P} \restriction B$ and even $\mathbb{P} \restriction A <_{\circ} \mathbb{P} \restriction B$. More explicitly, any $p \in \mathbb{P} \restriction B$ has a canonical reduction $p_0 = p_0(p, A, B) \in \mathbb{P} \restriction A$ such that*

- (i) $\text{dom}(p_0) = \text{dom}(p) \cap A$;
- (ii) $s_x^{p_0} = s_x^p$ for all $x \in \text{dom}(p_0) \cap L_{\text{Hech}}$ and $p_0(x) = p(x)$ for all $x \in \text{dom}(p_0) \cap L_{\text{mad}}$ (in particular, $n^{p_0} = n^p$),

and such that, whenever $D \in \mathcal{I}$, $B, C \subseteq D$, C closed, $C \cap B = A$, and $q_0 \in \mathbb{P} \restriction C$ extends p_0 , then there is $q \in \mathbb{P} \restriction D$ extending both q_0 and p .

Note that we do not require $p \leq_{\mathbb{P} \restriction B} p_0$.

Proof. By recursion-induction on α , simultaneously for all templates (L, \mathcal{I}) ,

- we prove that $\mathbb{P} \restriction B$ is indeed a p.o. (i.e., transitivity holds) for all $B \in \mathcal{I}$ with $\text{Dp}(B) = \alpha$;
- we prove $\mathbb{P} \restriction A \subset \mathbb{P} \restriction B$ for all $B \in \mathcal{I}$ with $\text{Dp}(B) = \alpha$ and all closed $A \subset B$;
- we construct $p_0 = p_0(p, A, B)$ satisfying (i) and (ii) for all $B \in \mathcal{I}$ with $\text{Dp}(B) = \alpha$, all closed $A \subset B$ and all $p \in \mathbb{P} \restriction B$;
- we prove that for all $B, D \in \mathcal{I}$, $B, C \subseteq D$, C closed, with $\text{Dp}(D) = \alpha$ and all $p \in \mathbb{P} \restriction B$, letting $A = C \cap B$ and $p_0 = p_0(p, A, B)$ (which has been constructed either at stage α or at an earlier stage), there is $q \in \mathbb{P} \restriction D$ as required.

Notice that if $\text{Dp}(B) = \alpha$, $D = B$ and $C = A$, then the latter indeed shows $\mathbb{P} \restriction A <_{\circ} \mathbb{P} \restriction B$.

The case $\alpha = 0$ is trivial. So assume $\alpha > 0$ and $\text{Dp}(B) = \alpha$. We first check transitivity of $\mathbb{P} \restriction B$. Assume that $r \leq_{\mathbb{P} \restriction B} q \leq_{\mathbb{P} \restriction B} p$. Then clearly $\text{dom}(r) \supseteq \text{dom}(p)$, $n^r \geq n^p$, $p(z) \subseteq r(z)$ for all $z \in \text{dom}(p) \cap L_{\text{mad}}$, and $|\{z \in \text{dom}(p) \cap L_{\text{mad}}; r(z)(i) = 1\}| \leq 1$ for all $i \in n^r \setminus n^p$. Let $z = \max(\text{dom}(r) \cap L_{\text{Hech}})$, $y = \max(\text{dom}(q) \cap L_{\text{Hech}})$ and $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$. Then there are $A_0, A_1 \in \mathcal{I} \cap \mathcal{P}(B \cap L_z)$

such that $p \restriction (B \cap L_z), q \restriction (B \cap L_z) \in \mathbb{P} \restriction A_0$, $q \restriction (B \cap L_z) \leq_{\mathbb{P} \restriction A_0} p \restriction (B \cap L_z)$, and $q \restriction (B \cap L_z), r \restriction (B \cap L_z) \in \mathbb{P} \restriction A_1$, $r \restriction (B \cap L_z) \leq_{\mathbb{P} \restriction A_1} q \restriction (B \cap L_z)$. Furthermore, \dot{f}_z^r is a $\mathbb{P} \restriction A_1$ -name and, in case $y = z$, \dot{f}_y^q is both a $\mathbb{P} \restriction A_1$ -name and a $\mathbb{P} \restriction A_0$ -name as well as, in case $x = y = z$, \dot{f}_x^p is a $\mathbb{P} \restriction A_0$ -name. Let $A = A_0 \cup A_1$. Then $A \in \mathcal{I} \cap \mathcal{P}(B \cap L_z)$ so that $\text{Dp}(A) < \text{Dp}(B)$, and we know by the induction hypothesis that $\mathbb{P} \restriction A_i < \circ \mathbb{P} \restriction A$ for $i = 0, 1$. Therefore, $p \restriction (B \cap L_z), q \restriction (B \cap L_z), r \restriction (B \cap L_z) \in \mathbb{P} \restriction A$ and $r \restriction (B \cap L_z) \leq_{\mathbb{P} \restriction A} q \restriction (B \cap L_z) \leq_{\mathbb{P} \restriction A} p \restriction (B \cap L_z)$. Moreover, \dot{f}_z^r is a $\mathbb{P} \restriction A$ -name and, in case $y = z$, \dot{f}_y^q is a $\mathbb{P} \restriction A$ -name and $r \restriction (B \cap L_z) \Vdash_{\mathbb{P} \restriction A} \dot{f}_y^q \leq \dot{f}_z^r$ as well as, in case $x = y = z$, \dot{f}_x^p is a $\mathbb{P} \restriction A$ -name and $q \restriction (B \cap L_z) \Vdash_{\mathbb{P} \restriction A} \dot{f}_x^p \leq \dot{f}_y^q$. Taken together, this shows that $r \leq_{\mathbb{P} \restriction B} p$ as required.

Now let $A \subset B \in \mathcal{I}$, A closed, be given. Assume $r \in \mathbb{P} \restriction A$. Let $x = \max(\text{dom}(r) \cap L_{\text{Hech}})$. By definition of the iteration there is $\bar{A} \in (\mathcal{I} \restriction A) \cap \mathcal{P}(L_x)$ such that $r \restriction (A \cap L_x) \in \mathbb{P} \restriction \bar{A}$ and \dot{f}_x^r is a $\mathbb{P} \restriction \bar{A}$ -name. There is $\bar{B} \in \mathcal{I} \restriction B \subseteq \mathcal{I}$ such that $\bar{A} = A \cap \bar{B}$. Clearly $x \notin \bar{B}$. By clause (4') in the definition of a template, we may therefore assume without loss of generality that $\bar{B} \subseteq L_x$. Thus $\bar{B} \subset B$ and $\text{Dp}(\bar{B}) < \text{Dp}(B) = \alpha$. By induction hypothesis, $\mathbb{P} \restriction \bar{A} \subset \mathbb{P} \restriction \bar{B}$ and $\mathbb{P} \restriction \bar{A} < \circ \mathbb{P} \restriction \bar{B}$. Therefore, \dot{f}_x^r is a $\mathbb{P} \restriction \bar{B}$ -name as well. $r \in \mathbb{P} \restriction B$ follows immediately. Hence $\mathbb{P} \restriction A \subset \mathbb{P} \restriction B$ as required.

Next assume also $p \in \mathbb{P} \restriction B$ is given. We construct $p_0 = p_0(p, A, B)$. Put $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$. By definition of the iteration there is $\bar{B} \in \mathcal{I} \cap \mathcal{P}(B \cap L_x)$ such that $\bar{p} = p \restriction (B \cap L_x) \in \mathbb{P} \restriction \bar{B}$ and \dot{f}_x^p is a $\mathbb{P} \restriction \bar{B}$ -name. Put $\bar{A} = A \cap \bar{B}$. Note that $\bar{A} \in \mathcal{I} \restriction A$. By induction hypothesis, \bar{p} has a reduction $\bar{p}_0 = p_0(\bar{p}, \bar{A}, \bar{B}) \in \mathbb{P} \restriction \bar{A}$ satisfying the barred version of the clauses of the main lemma. The definition of p_0 splits into two cases.

Case 1. $x \notin A$. Then $\text{dom}(p_0) = \text{dom}(p) \cap A$, $p_0 \restriction (A \cap L_x) = \bar{p}_0$ and $p_0(z) = p(z)$ for $z \in (\text{dom}(p) \cap A) \setminus L_x$. (Note that such z must belong to L_{mad} .)

Case 2. $x \in A$. Then let $\text{dom}(p_0) = \text{dom}(p) \cap A$, $p_0 \restriction (A \cap L_x) = \bar{p}_0$ and $p_0(z) = p(z)$ for $z \in (\text{dom}(p) \cap A \cap L_{\text{mad}}) \setminus L_x$. We know by induction hypothesis that $\mathbb{P} \restriction \bar{A} < \circ \mathbb{P} \restriction \bar{B}$. Therefore, there exists a *canonical projection* $\dot{f}_x^{p_0}$ to $\mathbb{P} \restriction \bar{A}$ of the $\mathbb{P} \restriction \bar{B}$ -name \dot{f}_x^p . Accordingly we let $p_0(x) = (s_x^p, \dot{f}_x^{p_0})$.

More explicitly, we do the following. For simplicity work with the cBa's $\mathbb{B}_{\bar{A}} = r.o.(\mathbb{P} \restriction \bar{A})$ and $\mathbb{B}_{\bar{B}} = r.o.(\mathbb{P} \restriction \bar{B})$ associated with \bar{A} and \bar{B} . We know by the induction hypothesis that $\mathbb{B}_{\bar{A}} < \circ \mathbb{B}_{\bar{B}}$. Note that $\bar{p} \leq \llbracket s_x^p \subseteq \dot{f}_x^p \rrbracket$. In $\mathbb{B}_{\bar{B}}$, for all $s \in \omega^{<\omega}$ with $s_x^p \subseteq s$, we let $b_s = \llbracket s \subseteq \dot{f}_x^p \rrbracket \cap \bar{p}$. So $b_{s_x^p} = \bar{p}$ and, for $n > |s_x^p|$, the b_s , $|s| = n$, are a maximal antichain below \bar{p} . Let a_s^* be the product (intersection) of \bar{p}_0 and the *projection* of b_s to $\mathbb{B}_{\bar{A}}$ (recall the projection of b_s to $\mathbb{B}_{\bar{A}}$ is the unique condition a such that $a \geq b_s$ and any extension of a in $\mathbb{B}_{\bar{A}}$ is compatible with b_s). In particular, $a_{s_x^p}^* = \bar{p}_0$ and $\sum \{a_s^*; |s| = n\} = \bar{p}_0$ for $n > |s_x^p|$. Define a_s by recursion on $n = |s|$ as follows. $a_{s_x^p} = a_{s_x^p}^* = \bar{p}_0$ and, for $n > |s_x^p|$, set $a_s = a_{s \restriction (n-1)} \cdot (a_s^* \setminus \sum_{j < s(n-1)} a_{s \restriction (n-1) \hat{\ } j}^*)$ (which is equal to $a_{s \restriction (n-1)} \cdot (a_s^* \setminus \sum_{j < s(n-1)} a_{s \restriction (n-1) \hat{\ } j})$). Then one can show by induction on $n > |s_x^p|$ that the a_s , $|s| = n$, are a maximal antichain below \bar{p}_0 . Therefore they canonically define a $\mathbb{P} \restriction \bar{A}$ -name $\dot{f}_x^{p_0}$ (that is, $a_s = \llbracket s \subseteq \dot{f}_x^{p_0} \rrbracket$) such that $\bar{p}_0 \Vdash_{\mathbb{P} \restriction \bar{A}} \dot{f}_x^{p_0} \in \omega^\omega$.

The main property of this name is that for all s , $a'_s = \sum \{a_{s'}; s' \leq s \text{ everywhere, } s_x^p \subseteq s, |s'| = |s|\}$ is a reduction (not necessarily the projection) of $b'_s = \sum \{b_{s'}; s' \leq s \text{ everywhere, } s_x^p \subseteq s, |s'| = |s|\}$. (This is so because $(a_s^*)' = \sum \{a_{s'}^*; s' \leq s$

everywhere, $s_x^p \subseteq s, |s'| = |s|$ is the product of \bar{p}_0 and the projection of b'_s and, by the definition of a_s , $a'_s \leq (a_s^*)'$ is trivial.)

This completes the definition of p_0 . Clauses (i) and (ii) are trivially satisfied in each of the two cases.

Now assume $B, D \in \mathcal{I}$, $B, C \subseteq D$, C closed, are such that $\text{Dp}(D) = \alpha$, $A = C \cap B$, $p \in \mathbb{P} \restriction B$ and $p_0 = p_0(p, A, B)$. Let $q_0 \leq_{\mathbb{P} \restriction C} p_0$, $q_0 \in \mathbb{P} \restriction C$. We need to construct q . $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$, \bar{B} , \bar{A} , \bar{p} and \bar{p}_0 are as in the previous construction.

Case 1. $x \notin A$. So $x \notin C$. Let $y = \max\{z < x; z \in \text{dom}(q_0) \cap L_{\text{Hech}}\}$. We can find $\bar{E} \in \mathcal{I} \restriction C \cap \mathcal{P}(L_y)$ such that $q_0 \restriction (C \cap L_y) \in \mathbb{P} \restriction \bar{E}$ and $\dot{f}_y^{q_0}$ is a $\mathbb{P} \restriction \bar{E}$ -name. There is $\bar{F} \in \mathcal{I} \restriction D \subseteq \mathcal{I}$ such that $\bar{E} = \bar{F} \cap C$. Since $y \notin \bar{E}$ and $y \in C$, $y \notin \bar{F}$ follows. By clause (4') in the definition of a template, without loss of generality, $\bar{F} \subseteq L_y$. By clause (3') in the definition of a template find $\bar{G} \in (\mathcal{I} \restriction D) \cap \mathcal{P}(L_x) \subseteq \mathcal{I}$ containing y . Let $\bar{D} = \bar{G} \cup \bar{F} \cup \bar{B} \subseteq L_x$ and $\bar{C} = (\bar{G} \cap C) \cup \bar{E} \cup \bar{A} \subseteq L_x$. By clause (2) in the definition of a template, $\bar{D} \in \mathcal{I} \restriction D \subseteq \mathcal{I}$ and $\bar{C} \in \mathcal{I} \restriction C$. Since $\text{Dp}_{\mathcal{I} \restriction C}(\bar{E}) \leq \text{Dp}_{\mathcal{I} \restriction C}(\bar{C}) \leq \text{Dp}_{\mathcal{I}}(\bar{D}) < \alpha$, $\bar{q}_0 = q_0 \restriction (C \cap L_x) \in \mathbb{P} \restriction \bar{C}$ by the induction hypothesis. $\bar{q}_0 \leq_{\mathbb{P} \restriction \bar{C}} \bar{p}_0$ and $\bar{C} \cap \bar{B} = \bar{A}$ are immediate. By the inductive assumption for the barred version, there is $\bar{q} \in \mathbb{P} \restriction \bar{D}$ extending both \bar{q}_0 and \bar{p} .

We define q such that

- $\text{dom}(q) = \text{dom}(\bar{q}) \cup \text{dom}(p) \cup \text{dom}(q_0)$, $n^q = n^{\bar{q}}$,
- $q \restriction (D \cap L_x) = \bar{q}$,
- $q(y) = q_0(y)$ for all $y \in (\text{dom}(q_0) \setminus L_x) \cap L_{\text{Hech}}$,
- $q(x) = p(x)$,
- $q_0(z) \subseteq q(z)$, $q(z)(i) = 0$, for $z \in (\text{dom}(q_0) \setminus L_x) \cap L_{\text{mad}}$ and $i \in n^q \setminus n^{q_0}$,
- $p(z) \subseteq q(z)$, $q(z)(i) = 0$, for $z \in (\text{dom}(p) \setminus (L_x \cup \text{dom}(q_0))) \cap L_{\text{mad}}$ and $i \in n^q \setminus n^p$.

It is straightforward to check that $q \in \mathbb{P} \restriction D$ and $q \leq_{\mathbb{P} \restriction D} q_0$. So let us argue that $q \leq_{\mathbb{P} \restriction D} p$ as well. Clearly $n^q \geq n^p$. We need to show that $p(z) \subseteq q(z)$ for all $z \in \text{dom}(p) \cap L_{\text{mad}}$. This is obvious for $z < x$ because $\bar{q} \leq_{\mathbb{P} \restriction \bar{D}} \bar{p}$. It is immediate by definition for $z > x$ belonging to $\text{dom}(p) \setminus \text{dom}(q_0)$. So assume $z > x$, $z \in \text{dom}(p) \cap \text{dom}(q_0)$. Then $p(z) = p_0(z) \subseteq q_0(z) \subseteq q(z)$, as required. Next fix $i \in n^q \setminus n^p$. We need to check that there is at most one $z \in \text{dom}(p) \cap L_{\text{mad}}$ with $q(z)(i) = 1$. By way of contradiction assume this is true for two distinct $z_0 < z_1$. By construction we must have $i \in n^{q_0}$, $x < z_1$ and $z_1 \in \text{dom}(q_0) \cap \text{dom}(p)$. Hence $z_1 \in A$. Therefore z_0 must belong to A as well because A is closed. Thus both z_0 and z_1 belong to $\text{dom}(p_0)$. This means that $q(z_j)(i) = q_0(z_j)(i) = 1$ for $j = 0, 1$, which contradicts $q_0 \leq_{\mathbb{P} \restriction C} p_0$, and we are done.

Case 2. $x \in A$. So $x \in C$. Find $\bar{C} \in \mathcal{I} \restriction C \cap \mathcal{P}(L_x)$ such that $\bar{q}_0 = q_0 \restriction (C \cap L_x) \in \mathbb{P} \restriction \bar{C}$ and $\dot{f}_x^{q_0}$ is a $\mathbb{P} \restriction \bar{C}$ -name. Without loss of generality, $\bar{A} \subseteq \bar{C}$. $\bar{C} \cap \bar{B} = \bar{A}$ is immediate. There is $\bar{D} \in \mathcal{I} \restriction D \subseteq \mathcal{I}$ such that $\bar{C} = \bar{D} \cap C$. Since $x \notin \bar{C}$, we get $x \notin \bar{D}$. By (4'), without loss of generality, $\bar{D} \subseteq L_x$. We may also assume $\bar{B} \subseteq \bar{D}$. Since $\text{Dp}_{\mathcal{I}}(\bar{D}) < \text{Dp}_{\mathcal{I}}(D) = \alpha$, we can freely use the induction hypothesis when dealing with \bar{A} , \bar{B} , \bar{C} , and \bar{D} . In particular, $\bar{q}_0 \leq_{\mathbb{P} \restriction \bar{C}} \bar{p}_0$.

Now note that we have $s_x^p = s_x^{p_0} \subseteq s_x^{q_0}$ and $\bar{q}_0 \Vdash_{\mathbb{P} \restriction \bar{C}} s_x^{q_0} \subseteq \dot{f}_x^{q_0}$. Let $m = |s_x^{q_0}|$. Since also $\bar{q}_0 \Vdash_{\mathbb{P} \restriction \bar{C}} \dot{f}_x^{q_0} \geq \dot{f}_x^{p_0}$ (everywhere), we see that $\bar{q}_0 \Vdash_{\mathbb{P} \restriction \bar{C}} \dot{f}_x^{p_0} \restriction m \leq s_x^{q_0}$. Hence we get $a := a'_{s_x^{q_0}} = [\dot{f}_x^{p_0} \restriction m \leq s_x^{q_0}] \geq_{\mathbb{P} \restriction \bar{A}} \bar{p}_0^*$ where we let $\bar{p}_0^* = p_0(\bar{q}_0, \bar{A}, \bar{C})$ (the canonical reduction of \bar{q}_0 to $\mathbb{P} \restriction \bar{A}$; note here that $\text{Dp}_{\mathcal{I} \restriction \bar{C}}(\bar{C}) \leq \text{Dp}_{\mathcal{I}}(\bar{D}) < \alpha$, so

that \bar{p}_0^* indeed has been defined already). However, by construction, $a \leq_{\mathbb{P} \restriction \bar{A}} \bar{p}_0$ is nothing but a reduction of $b := b'_{s_{x_0}^q} = \llbracket \dot{f}_x^p \restriction m \leq s_x^{q_0} \rrbracket \cap \bar{p} \in \mathbb{B}_{\bar{B}}$ to $\mathbb{B}_{\bar{A}}$. So there is $\bar{p}^+ \in \mathbb{P} \restriction \bar{B}$ such that $\bar{p}^+ \leq_{\mathbb{P} \restriction \bar{B}} \bar{p}_0^*$ and $\bar{p}^+ \leq_{\mathbb{P} \restriction \bar{B}} b$ (so that, in particular, $\bar{p}^+ \leq_{\mathbb{P} \restriction \bar{B}} \bar{p}$ and $\bar{p}^+ \Vdash_{\mathbb{P} \restriction \bar{B}} \dot{f}_x^p \restriction m \leq s_x^{q_0}$). Let $\bar{p}_0^+ = p_0(\bar{p}^+, \bar{A}, \bar{B})$ be the canonical reduction of \bar{p}^+ to $\mathbb{P} \restriction \bar{A}$. Then $\bar{p}_0^+ \leq_{\mathbb{P} \restriction \bar{A}} \bar{p}_0^*$. Therefore \bar{p}_0^+ and \bar{q}_0 have a common extension \bar{q}_0^+ in $\mathbb{P} \restriction \bar{C}$. By inductive assumption for the barred $+$ -version, there is $\bar{q}^+ \in \mathbb{P} \restriction \bar{D}$ extending both \bar{p}^+ and \bar{q}_0^+ .

We define q such that

- $\text{dom}(q) = \text{dom}(\bar{q}^+) \cup \text{dom}(p) \cup \text{dom}(q_0)$, $n^q = n^{\bar{q}^+}$,
- $q \restriction (D \cap L_x) = \bar{q}^+$,
- $q(y) = q_0(y)$ for all $y \in \text{dom}(q_0) \cap L_{\text{Hech}}$ with $y > x$,
- $s_x^q = s_x^{q_0}$ and \dot{f}_x^q is a $\mathbb{P} \restriction \bar{D}$ -name such that $\bar{q}^+ \Vdash_{\mathbb{P} \restriction \bar{D}} \dot{f}_x^q = \max\{\dot{f}_x^{q_0}, \dot{f}_x^p\}$,
- $q(z) \subseteq q(z)$, $q(z)(i) = 0$, for $z \in (\text{dom}(q_0) \setminus L_x) \cap L_{\text{mad}}$ and $i \in n^q \setminus n^{q_0}$,
- $p(z) \subseteq q(z)$, $q(z)(i) = 0$, for $z \in (\text{dom}(p) \setminus (L_x \cup \text{dom}(q_0))) \cap L_{\text{mad}}$ and $i \in n^q \setminus n^p$.

To see that $q \in \mathbb{P} \restriction D$, note that $\bar{q}^+ \Vdash_{\mathbb{P} \restriction \bar{D}} s_x^{q_0} \subseteq \dot{f}_x^q$ by construction. It is then straightforward to check that $q \leq_{\mathbb{P} \restriction D} q_0, p$. In fact, for $q \leq_{\mathbb{P} \restriction D} p$ we argue as in Case 1 above. \square

Note that, as an immediate consequence of Main Lemma 1.1, we get that for arbitrary closed $A \subseteq B \subseteq L$, $\mathbb{P} \restriction (A, \mathcal{I} \restriction A)$ completely embeds into $\mathbb{P} \restriction (B, \mathcal{I} \restriction B)$.

Lemma 1.2 (Chain condition). *Let $A \in \mathcal{I}$. Any uncountable $K \subseteq \mathbb{P} \restriction A$ has an uncountable centered subset.*

Proof. By a standard Δ -system argument, it suffices to show that if $p, q \in \mathbb{P} \restriction A$, $n^p = n^q$, $s_x^p = s_x^q$ for all $x \in \text{dom}(p) \cap \text{dom}(q) \cap L_{\text{Hech}}$, and $p(x) = q(x)$ for all $x \in \text{dom}(p) \cap \text{dom}(q) \cap L_{\text{mad}}$, then there is a common extension r with $\text{dom}(r) = \text{dom}(p) \cup \text{dom}(q)$, $n^r = n^p = n^q$,

$$s_x^r = \begin{cases} s_x^p & \text{if } x \in \text{dom}(p) \cap L_{\text{Hech}}, \\ s_x^q & \text{if } x \in \text{dom}(q) \cap L_{\text{Hech}} \end{cases}$$

and

$$r(x) = \begin{cases} p(x) & \text{if } x \in \text{dom}(p) \cap L_{\text{mad}}, \\ q(x) & \text{if } x \in \text{dom}(q) \cap L_{\text{mad}}. \end{cases}$$

We do this by induction on $\text{Dp}(A)$.

The case $\text{Dp}(A) = 0$ is trivial. So assume $\text{Dp}(A) > 0$. First assume $x = \max(\text{dom}(p) \cap L_{\text{Hech}}) < y = \max(\text{dom}(q) \cap L_{\text{Hech}})$. Then there is $B \in \mathcal{I} \cap \mathcal{P}(A \cap L_y)$ such that $p \restriction L_y, q \restriction L_y \in \mathbb{P} \restriction B$, and \dot{f}_y^q is a $\mathbb{P} \restriction B$ -name. By the induction hypothesis, we get the required $r \restriction L_y \leq_{\mathbb{P} \restriction B} p \restriction L_y, q \restriction L_y$. Let $r(y) = q(y)$ and let $r(z) = p(z)$ for $z \in \text{dom}(p) \cap L_{\text{mad}}$, $z > y$, and $r(z) = q(z)$ for $z \in (\text{dom}(q) \setminus \text{dom}(p)) \cap L_{\text{mad}}$, $z > y$.

Next assume $x = \max(\text{dom}(p) \cap L_{\text{Hech}}) = \max(\text{dom}(q) \cap L_{\text{Hech}})$. Again there is $B \in \mathcal{I} \cap \mathcal{P}(A \cap L_x)$ such that $p \restriction L_x, q \restriction L_x \in \mathbb{P} \restriction B$, and \dot{f}_x^p, \dot{f}_x^q are $\mathbb{P} \restriction B$ -names. Again we get $r \restriction L_x$. Let $s_x^r = s_x^p = s_x^q$ and \dot{f}_x^r be such that $r \restriction L_x \Vdash_{\mathbb{P} \restriction B} \dot{f}_x^r = \max\{\dot{f}_x^p, \dot{f}_x^q\}$. Also let $r(z) = p(z)$ for $z \in \text{dom}(p) \cap L_{\text{mad}}$, $z > y$, and $r(z) = q(z)$ for $z \in (\text{dom}(q) \setminus \text{dom}(p)) \cap L_{\text{mad}}$, $z > y$. \square

Lemma 1.3 (Embedding Hechler forcing). *Let $x \in L_{\text{Hech}}$ and $A \in \mathcal{I} \cap \mathcal{P}(L_x)$. Then the two-step iteration $\mathbb{P} \restriction A \star \dot{\mathbb{D}}_x$ that canonically adds a Hechler-generic in coordinate x over the generic extension via $\mathbb{P} \restriction A$ completely embeds into $\mathbb{P} \restriction L$.*

Proof. Let $B = \text{cl}(A \cup \{x\})$. $\mathbb{P} \restriction B$ embeds into $\mathbb{P} \restriction L$ by Main Lemma 1.1. So it suffices to show $\mathbb{P} \restriction A \star \dot{\mathbb{D}}_x < \circ \mathbb{P} \restriction B$. This does not follow from (the statement of) Lemma 1.1 because $A \cup \{x\}$ need not be closed, but it is relatively straightforward from the proof of 1.1.

More explicitly, given $p \in \mathbb{P} \restriction B$, there is $\bar{B} \in \mathcal{I} \restriction B \cap \mathcal{P}(L_x)$ such that $\bar{p} = p \restriction L_x \in \mathbb{P} \restriction \bar{B}$ and \dot{f}_x^p is a $\mathbb{P} \restriction \bar{B}$ -name. Without loss of generality, $A \subseteq \bar{B}$. By 1.1, $\mathbb{P} \restriction A < \circ \mathbb{P} \restriction \bar{B}$. Therefore, \bar{p} has a canonical reduction $\bar{p}_0 \in \mathbb{P} \restriction A$. As in Case 2 of the proof of 1.1, there is a canonical projection $\dot{f}_x^{p_0}$ to $\mathbb{P} \restriction A$ of \dot{f}_x^p . Define $p_0 \in \mathbb{P} \restriction A \star \dot{\mathbb{D}}_x$ by $p_0 \restriction A = \bar{p}_0$ and $p_0(x) = (s_x^p, \dot{f}_x^{p_0})$. As in Case 2 of the proof of 1.1, argue that any $q_0 \in \mathbb{P} \restriction A \star \dot{\mathbb{D}}_x$ extending p_0 is compatible with p . \square

This may badly fail in case $A \notin \mathcal{I}$ because then $\mathbb{P} \restriction A \star \dot{\mathbb{D}}_x$ need not embed into $\mathbb{P} \restriction B$.

Lemma 1.4 (Names for reals). *Assume $p \in \mathbb{P} \restriction L$ and \dot{f} is a $\mathbb{P} \restriction L$ -name for a real. Then there is $A \subseteq L$ countable such that, letting $B = \text{cl}(A)$, $p \in \mathbb{P} \restriction B$, and \dot{f} is a $\mathbb{P} \restriction B$ -name.*

Proof. The proof proceeds by simultaneous induction on $\text{Dp}(L)$. Without loss of generality, $\text{Dp}(L) > 0$.

Assume first $p \in \mathbb{P} \restriction L$. Let $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$. There is $C \in \mathcal{I} \cap \mathcal{P}(L_x)$ such that $p \restriction L_x \in \mathbb{P} \restriction C$ and \dot{f}_x^p is a $\mathbb{P} \restriction C$ -name. By induction hypothesis, there is $A_0 \subseteq C$ countable such that $p \restriction L_x \in \mathbb{P} \restriction \text{cl}(A_0)$ and \dot{f}_x^p is a $\mathbb{P} \restriction \text{cl}(A_0)$ -name. Then $p \in \mathbb{P} \restriction B$ where $B = \text{cl}(A)$, $A = A_0 \cup \text{dom}(p)$.

Assume now that \dot{f} is a $\mathbb{P} \restriction L$ -name. By ccc-ness (Lemma 1.2), there are $\{p_{n,i}; i, n \in \omega\} \subseteq \mathbb{P} \restriction L$ and $\{k_{n,i} \in \omega; i, n \in \omega\}$ such that

- $p_{n,i} \Vdash_{\mathbb{P} \restriction L} \dot{f}(n) = k_{n,i}$,
- $\{p_{n,i}; i \in \omega\}$ is a maximal antichain in $\mathbb{P} \restriction L$ for all $n \in \omega$.

By the previous paragraph, we can find countable sets $A_{n,i}$ such that $p_{n,i} \in \mathbb{P} \restriction \text{cl}(A_{n,i})$. Put $A = \bigcup_{i,n} A_{n,i}$, $B = \text{cl}(A)$. Since $\mathbb{P} \restriction B < \circ \mathbb{P} \restriction L$ (Lemma 1.1), we can construe \dot{f} as a $\mathbb{P} \restriction B$ -name. \square

Assume (L, \mathcal{I}) and (L, \mathcal{J}) are templates and $\mathcal{I} \subseteq \mathcal{J}$. We say that \mathcal{I} is *cofinal* in \mathcal{J} if for all $x \in L_{\text{Hech}}$ and all $A \in \mathcal{J} \cap \mathcal{P}(L_x)$ there is $B \in \mathcal{I} \cap \mathcal{P}(L_x)$ containing A . The following is, in a sense, a triviality.

Lemma 1.5 (Cofinal subtemplates). *If \mathcal{I} is cofinal in \mathcal{J} , then $\mathbb{P} \restriction (L, \mathcal{I})$ is forcing equivalent to $\mathbb{P} \restriction (L, \mathcal{J})$.*

Proof. By induction on $\text{Dp}(L)$ (in the sense of \mathcal{I}), we argue that conditions in $\mathbb{P} \restriction (L, \mathcal{I})$ and conditions in $\mathbb{P} \restriction (L, \mathcal{J})$ can be canonically identified so as to yield forcing equivalence. Without loss of generality, $\text{Dp}(L) > 0$.

Let $p \in \mathbb{P} \restriction (L, \mathcal{J})$. Put $x = \max(\text{dom}(p) \cap L_{\text{Hech}})$. There is $A \in \mathcal{J} \cap \mathcal{P}(L_x)$ such that $\bar{p} = p \restriction L_x \in \mathbb{P} \restriction (A, \mathcal{J})$ and \dot{f}_x^p is a $\mathbb{P} \restriction (A, \mathcal{J})$ -name. Since \mathcal{I} is cofinal in \mathcal{J} , there is $B \in \mathcal{I} \cap \mathcal{P}(L_x)$ such that $A \subseteq B$. By Main Lemma 1.1, we know that $\mathbb{P} \restriction (A, \mathcal{J}) < \circ \mathbb{P} \restriction (B, \mathcal{J})$ and, by the induction hypothesis, $\mathbb{P} \restriction (B, \mathcal{J})$ and $\mathbb{P} \restriction (B, \mathcal{I})$ are forcing equivalent. Therefore, we may construe \bar{p} as a condition in $\mathbb{P} \restriction (B, \mathcal{I})$ and \dot{f}_x^p as a $\mathbb{P} \restriction (B, \mathcal{I})$ -name. Thus $p \in \mathbb{P} \restriction (L, \mathcal{I})$. It is straightforward to verify that this identification induces forcing equivalence. \square

Proposition 1.6 (Adjoining a scale). *Assume μ is regular uncountable, $\mu \subseteq L_{\text{Hech}}$ is cofinal in L , and $L_\alpha \in \mathcal{I}$ for all $\alpha < \mu$. Then $\mathbb{P} \restriction L$ forces $\mathfrak{b} = \mathfrak{d} = \mu$ (i.e., there is a μ -scale).*

Proof. For each $\alpha < \mu$, let \dot{f}_α be the name for the Hechler-generic adjoined in coordinate α of the iteration (see Lemma 1.3). By construction, the \dot{f}_α are forced to be well-ordered by \leq^* . Let \dot{g} be a $\mathbb{P} \restriction L$ -name for a real. By Lemma 1.4, there is $A \subseteq L$ countable such that \dot{g} is a $\mathbb{P} \restriction \text{cl}(A)$ -name. Since μ is regular uncountable and cofinal in L , there is $\alpha < \mu$ such that $\text{cl}(A) \subseteq L_\alpha$. Since $L_\alpha \in \mathcal{I}$, \dot{f}_α is forced to dominate the reals in the generic extension via $\mathbb{P} \restriction L_\alpha$ and, a fortiori, it will dominate \dot{g} . \square

Proposition 1.7 (Adjoining a mad family). *Assume L has uncountable cofinality and L_{mad} is cofinal in L . Then $\mathbb{P} \restriction L$ canonically adjoins a mad family of size $|L_{\text{mad}}|$.*

Proof. Let G be $\mathbb{P} \restriction L$ -generic over the ground model. For $x \in L_{\text{mad}}$ define $Y_x = \{n \in \omega; p(x)(n) = 1 \text{ for some } p \in G\}$. Let $\mathcal{A} = \{Y_x; x \in L_{\text{mad}}\}$. By definition of the p.o., \mathcal{A} is an almost-disjoint family. We need to check maximality. So let \dot{Z} be a $\mathbb{P} \restriction L$ -name for an infinite subset of ω and assume by way of contradiction that p forces that \dot{Z} is almost disjoint from all \dot{Y}_x . By Lemma 1.4 there is a countable set A such that $p \in \mathbb{P} \restriction \text{cl}(A)$ and \dot{Z} is a $\mathbb{P} \restriction \text{cl}(A)$ -name. Since L has uncountable cofinality and L_{mad} is cofinal in L , there is $x \in L_{\text{mad}}$ such that $\text{cl}(A) \subseteq L_x$. By Main Lemma 1.1 we know that $\mathbb{P} \restriction \text{cl}(A) < \mathbb{P} \restriction L_x < \mathbb{P} \restriction L$.

Find k_0 and $p_0 \leq_{\mathbb{P} \restriction L} p$ such that

$$p_0 \Vdash_{\mathbb{P} \restriction L} \dot{Z} \cap \dot{Y}_x \subseteq k_0.$$

Put $\bar{p}_0 = p_0 \restriction L_x$. Clearly any \dot{Y}_y , $y \in \text{dom}(\bar{p}_0) \cap L_{\text{mad}}$, is a $\mathbb{P} \restriction L_x$ -name. So we can find $k_1 \geq k_0$ and $\bar{p}_1 \leq_{\mathbb{P} \restriction L_x} \bar{p}_0$ such that

$$\bar{p}_1 \Vdash_{\mathbb{P} \restriction L_x} \dot{Z} \cap \dot{Y}_y \subseteq k_1$$

for all $y \in \text{dom}(\bar{p}_0) \cap L_{\text{mad}}$. Since \dot{Z} is forced to be infinite, we can find $\bar{p}_2 \leq_{\mathbb{P} \restriction L_x} \bar{p}_1$ and $i_0 \geq k_1$ such that $\bar{p}_2 \Vdash_{\mathbb{P} \restriction L_x} i_0 \in \dot{Z}$. Without loss of generality, $n^{\bar{p}_2} > i_0$. Then we must necessarily have $\bar{p}_2(y)(i_0) = 0$ for all $y \in \text{dom}(\bar{p}_0) \cap L_{\text{mad}}$.

Define a condition p_2 by

- $\text{dom}(p_2) = \text{dom}(\bar{p}_2) \cup \text{dom}(p_0)$, $n^{p_2} = n^{\bar{p}_2}$,
- $p_2 \restriction L_x = \bar{p}_2$,
- $p_2(z) = p_0(z)$ for all $z \in L_{\text{Hech}} \cap \text{dom}(p_0)$, $z > x$,
- $p_2(z) \supset p_0(z)$, $p_2(z)(i) = 0$ for all i with $n^{p_0} \leq i < n^{p_2}$ and all $z \in L_{\text{mad}} \cap \text{dom}(p_0)$, $z > x$,
- $p_2(x) \supset p_0(x)$,

$$p_2(x)(i) = \begin{cases} 1 & \text{if } i = i_0, \\ 0 & \text{for all } i \neq i_0 \text{ with } n^{p_0} \leq i < n^{p_2}. \end{cases}$$

It is straightforward to verify that $p_2 \in \mathbb{P} \restriction L$ and that $p_2 \leq_{\mathbb{P} \restriction L} p_0$. Since

$$p_2 \Vdash_{\mathbb{P} \restriction L} i_0 \in \dot{Z} \cap \dot{Y}_x,$$

we have a contradiction. \square

2. BUILDING A TEMPLATE FOR ADJOINING A MAD FAMILY

For simplicity assume CH for the remainder of the paper.

Assume $\lambda_0 \geq \aleph_2$ is regular, and $\lambda > \lambda_0$ is a singular cardinal of countable cofinality, say $\lambda = \bigcup_n \lambda_n$, the λ_n being regular, equal to $\lambda_n^{\aleph_0}$, and strictly increasing. Also suppose $\kappa^{\aleph_0} < \lambda_n$ for $\kappa < \lambda_n$. As usual, μ^* denotes (a disjoint copy of) μ with the reverse ordering. Elements of μ will be called *positive*, and members of μ^* are *negative*. For each n choose a partition $\lambda_n^* = \bigcup_{\alpha < \omega_1} S_n^\alpha$ such that each S_n^α is co-initial in λ_n^* . Also assume $S_n^\alpha \cap \lambda_m^* = S_m^\alpha$ for $m < n$.

The following definition is motivated by Shelah's work [S2].

Definition (Template for adjoining a mad family). Define $L = L(\lambda)$ as follows. Elements of L are nonempty finite sequences x (i.e., $\text{dom}(x) \in \omega$) such that

- $x(0) \in \lambda_0$,
- $x(n) \in \lambda_n^* \cup \lambda_n$ for $0 < n < |x| - 1$, and
- in case $|x| \geq 2$, if $x(|x| - 2)$ is positive, then $x(|x| - 1) \in \lambda_{|x|-1}^* \cup \lambda$, and if $x(|x| - 2)$ is negative, then $x(|x| - 1) \in \lambda^* \cup \lambda_{|x|-1}$.

Say $x \in L_{\text{Hech}}$ if $|x| = 1$ or $x(|x| - 1) \in \lambda_{|x|-1}^* \cup \lambda_{|x|-1}$. Otherwise $x \in L_{\text{mad}}$. (This means that $x \in L_{\text{mad}}$ iff $|x| \geq 2$ and either $x(|x| - 2)$ is positive and $x(|x| - 1) \geq \lambda_{|x|-1}$ or $x(|x| - 2)$ is negative and $x(|x| - 1) \leq \lambda_{|x|-1}^*$.) Equip L with the following lexicographic-like ordering: $x < y$ iff

- either $x \subset y$ and $y(|x|)$ is positive,
- or $y \subset x$ and $x(|y|)$ is negative,
- or, letting $n := \min\{m : x(m) \neq y(m)\}$, either $x(n)$ is negative and $y(n)$ is positive, or both are positive and $x(n) <_\lambda y(n)$, or both are negative and $x(n) <_{\lambda^*} y(n)$ (i.e., there are $\alpha < \beta < \lambda$ such that $x(n) = \beta^* <_{\lambda^*} \alpha^* = y(n)$).

It is immediate that this is indeed a linear ordering.

We identify sequences of length one with their ranges so that λ_0 is a cofinal subset of L . Say $x \in L_{\text{Hech}}$ is *relevant* if $|x| \geq 3$ is odd, $x(n)$ is negative for odd n and positive for even n , $x(|x| - 1) < \omega_1$, and whenever $n < m$ are even such that $x(n), x(m) < \omega_1$, then there are $\beta < \alpha$ such that $x(n - 1) \in S_{n-1}^\alpha$ and $x(m - 1) \in S_{m-1}^\beta$. For relevant x , set $J_x = [x \restriction (|x| - 1), x)$, the interval of nodes between $x \restriction (|x| - 1)$ and x in the order of L . Notice that if $x < y$ are relevant, then either $J_x \cap J_y = \emptyset$ or $J_x \subset J_y$ (in which case we also have $|y| \leq |x|$, $x \restriction (|y| - 1) = y \restriction (|y| - 1)$ and $x(|y| - 1) \leq y(|y| - 1)$).

Define $\mathcal{I} = \mathcal{I}(\lambda)$ to be the collection of all finite unions of sets of the form

- L_α for $\alpha \leq \lambda_0$,
- $\text{cl}(J_x)$ for relevant x ,
- $\text{cl}(\{x\})$ for $x \in L_{\text{Hech}}$, and
- $L_x \cap L_{\text{mad}}$ for $x \in L_{\text{Hech}}$.

So $L(\lambda)$ is a subtree of $(\lambda^* \cup \lambda)^{<\omega}$ (i.e., it is closed under taking initial segments). The nodes belonging to L_{mad} are exactly the terminal (= maximal) nodes of this tree. The point of the J_x is that we need “copies” of the large supports given by the L_α for isomorphism-of-names arguments. The S_n^α , then, are used to code the places where we put the J_x so that we basically get well-foundedness for free.

Lemma 2.1. (L, \mathcal{I}) is a template.

Proof. Clauses (1) and (6) in the definition of template are immediate, as is closure under finite unions. To see closure under finite intersections, it suffices to argue that the intersection of any two sets of the above form (i.e., L_α , $\text{cl}(J_x)$, $\text{cl}(\{x\})$, and $L_x \cap L_{\text{mad}}$) is again of this form. This, however, is straightforward so that (2) holds as well.

To prove (3'), let $x \in L_{\text{Hech}}$ and $y \in L_x$. In case $y \in L_{\text{Hech}}$, we have $y \in \text{cl}(\{y\}) \subseteq L_x$. If $y \in L_{\text{mad}}$, $y \in L_x \cap L_{\text{mad}} \subseteq L_x$.

For (4'), it suffices again to consider sets A from \mathcal{I} of the above form. Let $x \in L_{\text{Hech}} \setminus A$. Without loss of generality, $A \setminus L_x \neq \emptyset$. If A is of the form L_y , $\text{cl}(J_y)$, $\text{cl}(\{y\})$ or $L_y \cap L_{\text{mad}}$, then we must have $y > x$. $A = L_y$ is impossible and if $A = \text{cl}(J_y)$, then $x < y \upharpoonright (|y| - 1) = \min(J_y)$. So, in each of the possible cases, the intersection with L_x is $L_x \cap L_{\text{mad}}$.

We are left with showing well-foundedness (5'). Assume A_n , $n \in \omega$, is a decreasing chain from $\mathcal{I} \restriction L_{\text{Hech}}$. Let α_n be such that $L_{\alpha_n} \cap L_{\text{Hech}}$ occurs in A_n as a component. Choose α_{n_0} minimal among the α_n . Without loss of generality, $n_0 = 0$. Then all $L_{\alpha_n} \cap L_{\text{Hech}}$ are the same and it suffices to consider the J_x -components. Thus we may assume, without loss of generality, that $A_0 = J_{x_0} \cap L_{\text{Hech}}$, and there is a finitely-branching tree $T \subseteq \omega^{<\omega}$ such that $A_n = (\bigcup_{\sigma \in T \cap \omega^n} J_{x_n^\sigma} \cup F_n) \cap L_{\text{Hech}}$ where the $F_n \subseteq L_{\text{Hech}}$ are finite, and such that $\sigma \subseteq \tau$, $|\sigma| = n < |\tau| = m$, implies $J_{x_m^\tau} \subseteq J_{x_n^\sigma}$, and such that the $J_{x_n^\sigma}$, $\sigma \in T \cap \omega^n$, are pairwise disjoint. Now note that if $f \in [T]$ is a branch, then the sequence $\{x_n^{f \upharpoonright n}; n \in \omega\}$ must eventually stabilize. (First argue that if $|x_n^{f \upharpoonright n}| \rightarrow \infty$, then $\{\alpha; x_n^{f \upharpoonright n}(|x_n^{f \upharpoonright n}| - 2) \in S_n^\alpha \text{ for some } n\}$ would constitute a decreasing sequence of ordinals. Then notice that if $|x_n^{f \upharpoonright n}|$ is eventually constant, so is the decreasing sequence $x_n^{f \upharpoonright n}(|x_n^{f \upharpoonright n}| - 1)$.) Since T is a finitely-branching tree this means that the total number of the x_n^σ is finite which in turn entails that the sequence of the A_n eventually stabilizes. \square

Corollary 2.2 (Bounds for \mathfrak{a}). $\mathbb{P} \restriction L$ forces $\mathfrak{b} = \mathfrak{d} = \lambda_0$ and adjoins a mad family of size λ (so that $\lambda_0 \leq \mathfrak{a} \leq \lambda$).

Proof. This is immediate by construction of $\mathbb{P} \restriction L$ and Propositions 1.6 and 1.7. \square

3. KILLING MAD FAMILIES USING TEMPLATES

We are left with showing there is no mad family of size less than λ in the generic extension. As explained in the Introduction, this is an (albeit sophisticated) isomorphism-of-names argument. Isomorphisms of names canonically boil down to certain brands of partial isomorphisms between subsets of L , and we begin with their investigation.

Definition (Isomorphism). Let $A, B \subseteq L$ be countable trees.⁸ Call A and B *isomorphic* ($A \cong B$) iff there is a bijection $\phi = \phi_{A,B} : A \rightarrow B$ such that

- (a) $|\phi(x)| = |x|$,
- (b) $\phi(x) \upharpoonright n = \phi(x \upharpoonright n)$,
- (c) $x < y$ iff $\phi(x) < \phi(y)$,
- (d) $x(n)$ is positive iff $\phi(x)(n)$ is positive,
- (e) $x \in L_{\text{mad}}$ iff $\phi(x) \in L_{\text{mad}}$

for all $x, y \in A$ and all $n \in \omega$, and such that

⁸Recall A is a *tree* if it is closed under taking initial segments, i.e., given $x \in A$, we have $x \upharpoonright n \in A$ for all $n \in \omega$.

(f) $\mathcal{I} \restriction A$ is mapped to $\mathcal{I} \restriction B$ via ϕ .

Since the trace of \mathcal{I} on each countable set is countable, there are at most $2^{\aleph_0} = \aleph_1$ isomorphism types.

This, the strongest notion of “isomorphism” we shall consider, will be used in several pruning arguments below. However, for most purposes the following is sufficient.

Definition (Weak isomorphism). Let $A, B \subseteq L$ be arbitrary. We say that A and B are *weakly isomorphic* ($A \cong_{\text{weak}} B$) if (e) is satisfied and instead of clauses (c), (f) we have

(c') $x < y$ iff $\phi(x) < \phi(y)$ for all x, y such that there is $z \in L_{\text{Hech}} \cap A$ with $x \leq z \leq y$,

and

(f') ϕ maps a cofinal subset of $\mathcal{I} \restriction A$ to a cofinal subset of $\mathcal{I} \restriction B$,

respectively.

Lemma 3.1. *Let A and B be countable trees such that $L_{\text{Hech}} \cap A$ ($L_{\text{Hech}} \cap B$, respectively) is cofinal in A (in B , resp.). If $A \cong B$, as witnessed by ϕ , then there is ψ extending ϕ and witnessing that $\text{cl}(A) \cong_{\text{weak}} \text{cl}(B)$.*

Proof. Call a nonempty $X \subseteq \text{cl}(A) \cap L_{\text{mad}}$ *connected* if given $x < y$ from X , the interval $[x, y]$ is disjoint from $A \cap L_{\text{Hech}}$. A maximal connected set is called a *connected component*. Note every connected component has size λ (because L_{Hech} is cofinal in A) and $\text{cl}(A) \cap L_{\text{mad}}$ is a disjoint union of at most countably many connected components.

Given $x \in L_{\text{Hech}} \cap A$ with $x(|x| - 1)$ being positive, put $\text{Comp}_x = \{y \in L_{\text{mad}}; y < z \text{ for every } z \in A \cap L_{\text{Hech}} \text{ with } z \supseteq x \text{ and } y > z \text{ for every } z \in A \cap L_{\text{Hech}} \text{ such that } z \restriction |x| < x\}$. Clearly Comp_x is a connected component. Dually, define Comp_x for $x \in L_{\text{Hech}} \cap A$ with negative $x(|x| - 1)$ by interchanging $<$ and $>$.

For each $y \in \text{cl}(A) \cap L_{\text{mad}}$, there is $x \in L_{\text{Hech}} \cap A$ with $y \in \text{Comp}_x$. To see this, let $n < |y|$ be maximal such that $y \restriction n \in A$. Assume, without loss of generality, $y(n)$ is positive. Let $k \leq n$ be minimal such that all $y(i)$ for $k \leq i \leq n$ are positive. If possible choose m , $k \leq m \leq n$, and $x \in A \cap L_{\text{Hech}}$, $|x| = m + 1$, such that $x \restriction m = y \restriction m$, $x(m) > y(m)$ is minimal, and such that m is the maximal value for which such an x can be found. Then $y \in \text{Comp}_x$. If m and x cannot be found, we let $x = y \restriction k$ and check $y \in \text{Comp}_x$ (note that $x(|x| - 1) = y(k - 1)$ is negative in this case so that the second alternative of the definition of Comp applies). Therefore $\text{cl}(A) \cap L_{\text{mad}} = \bigcup_{x \in L_{\text{Hech}} \cap A} \text{Comp}_x$.

Also notice that for $x, x' \in L_{\text{Hech}} \cap A$, if $\text{Comp}_x = \text{Comp}_{x'}$, then $\text{Comp}_{\phi(x)} = \text{Comp}_{\phi(x')}$, and if $\text{Comp}_x \cap \text{Comp}_{x'} = \emptyset$, then $\text{Comp}_{\phi(x)} \cap \text{Comp}_{\phi(x')} = \emptyset$.

So we can simply extend ϕ to ψ by mapping Comp_x to $\text{Comp}_{\phi(x)}$ for all $x \in L_{\text{Hech}} \cap A$. Then (c') and (e) are immediate. To see (f'), note that, by definition of the template, sets in $\mathcal{I} \restriction \text{cl}(A)$ that are unions of sets from $\mathcal{I} \restriction A$ and of sets of the form $L_x \cap L_{\text{mad}}$ are cofinal in $\mathcal{I} \restriction \text{cl}(A)$. However, since ϕ identifies sets of $\mathcal{I} \restriction A$ and sets of $\mathcal{I} \restriction B$, ψ identifies sets of the latter kind. \square

Note that we did not use the full strength of our notion of isomorphism in the above proof. Clauses (c) and (f) could be replaced by (c') and (f') respectively. Furthermore, instead of dealing with trees A and B (and having (a), (b), and (d)),

it suffices that $\text{cl}(A) \cap L_{\text{mad}}$ is the union of the components Comp_x , $x \in L_{\text{Hech}} \cap A$, and similarly for B , and that extending ϕ by mapping Comp_x to $\text{Comp}_{\phi(x)}$ preserves (c').

Lemma 3.2. *If $A \cong_{\text{weak}} B$, then $\mathbb{P} \restriction A \cong \mathbb{P} \restriction B$.*

Proof. Notice that clauses (c), (e) and (f) are enough to guarantee that $\mathbb{P} \restriction A \cong \mathbb{P} \restriction B$. By Lemma 1.5, this is still true if (f) is replaced by (f'). Finally, by the way $\mathbb{P} \restriction A$ is defined recursively, interchanging elements of L_{mad} that belong to the same connected component of $A \cap L_{\text{mad}}$ does not affect the p.o.⁹ (because the interchanging map sends a cofinal subset of $\mathcal{I} \restriction A$ to a cofinal subset of $\mathcal{I} \restriction A$, see 1.5). □

Completion of the proof of the Main Theorem. Now assume \dot{A} is a name for an almost-disjoint family of size $< \lambda$, say \dot{A} is listed as $\{\dot{A}^\alpha; \alpha < \kappa\}$. Also assume \dot{A} is forced to have size at least λ_0 . Let $k < \omega$ be maximal such that $\kappa \geq \lambda_k$. Without loss of generality, $\kappa \geq \lambda_k \cdot 2$. We shall perform several standard pruning arguments, reordering the family of the \dot{A}^α so that the first λ_k many look very “similar”, that is, those \dot{A}^α that do not fit the pattern get removed to higher indices. This is why we stipulate $\kappa \geq \lambda_k \cdot 2$. Eventually, the first ω_1 many \dot{A}^α will suffice, and it is those that we use to create a new name \dot{A}^κ witnessing non-maximality.

For fixed α , find countable maximal antichains $\{p_{n,i}^\alpha; i \in \omega\} \subseteq \mathbb{P} \restriction L$, $n \in \omega$, and $\{k_{n,i}^\alpha \in 2; i, n \in \omega\}$ such that $p_{n,i}^\alpha \Vdash n \in \dot{A}^\alpha$ iff $k_{n,i}^\alpha = 1$ and $p_{n,i}^\alpha \Vdash n \notin \dot{A}^\alpha$ iff $k_{n,i}^\alpha = 0$. Let $B^\alpha = \bigcup \{\text{dom}(p_{n,i}^\alpha); i, n \in \omega\} \subseteq L$. B^α is at most countable. Without loss of generality, it is a tree. Let $C^\alpha = \text{cl}(B^\alpha)$. Put $B := \bigcup_\alpha B^\alpha$. So $|B| < \lambda_{k+1} < \lambda$. By CH and the Δ -system lemma we may assume, without loss of generality, that the $\{B^\alpha; \alpha < \lambda_k\}$ form a Δ -system, and that the bijection $\phi = \phi^{\alpha,\beta}$ (see above) sending B^α to B^β is an isomorphism fixing the root R of the system. Because there are only $\lambda_{k-1}^{\aleph_0} = \lambda_{k-1}$ many countable subsets of $L_{\text{Hech}} \cap (\lambda^* \cup \lambda)^k$, we may also assume that if $x \in B^\alpha \cap L_{\text{Hech}}$ and $|x| \leq k$, then $x \in R$. Also stipulate that there is some $\theta_0 < \omega_1$ such that whenever $\alpha < \lambda_k$, $x \in B^\alpha$, j odd and $x(j) \in \lambda_j^*$, then $x(j) \in S_j^\theta$ for some $\theta < \theta_0$. As explained above, ϕ canonically induces a weak isomorphism $\psi = \psi^{\alpha,\beta}$ between C^α and C^β (Lemma 3.1), which in turn yields an isomorphism $\chi = \chi^{\alpha,\beta}$ between $\mathbb{P} \restriction C^\alpha$ and $\mathbb{P} \restriction C^\beta$ (Lemma 3.2) both of which embed into $\mathbb{P} \restriction L$ (Main Lemma 1.1), as well as between $\mathbb{P} \restriction C^\alpha$ -names and $\mathbb{P} \restriction C^\beta$ -names. Furthermore, since connected components are homogeneous from the forcing point of view, since $C^\alpha \cap L_{\text{Hech}} = B^\alpha \cap L_{\text{Hech}}$ is countable, and since $C^\alpha \cap L_{\text{mad}}$ has only countably many connected components (see the proofs of Lemmas 3.1 and 3.2), it has, up to isomorphism, only $2^{\aleph_0} = \aleph_1$ many isomorphism types of names. (Of course, there are a total of λ^{\aleph_0} names.) Therefore, we may also suppose that χ identifies \dot{A}^α with \dot{A}^β , which means, more explicitly, that $k_{n,i} := k_{n,i}^\alpha = k_{n,i}^\beta$ and $\chi(p_{n,i}^\alpha) = p_{n,i}^\beta$.

Write $B^\alpha = \{x_s^\alpha; s \in T\}$ where $T \subseteq (\omega_1^* \cup \omega_1)^{<\omega}$ is the canonical tree isomorphic to any B^α . This means in particular that $\phi^{\alpha,\beta}(x_s^\alpha) = x_s^\beta$, that $|s| = |x_s^\alpha|$, and that $s(n)$ is positive iff $x_s^\alpha(n)$ is positive. Let $T_{\text{Hech}} = \{s \in T; x_s^\alpha \in L_{\text{Hech}}\}$ and $T_{\text{mad}} = \{s \in T; x_s^\alpha \in L_{\text{mad}}\}$. T_{Hech} is a subtree of T , while T_{mad} is a set of terminal nodes of T . Furthermore, let $S \subseteq T$ be the subtree of T corresponding to the root,

⁹That is, the order structure on connected components of $A \cap L_{\text{mad}}$ is irrelevant, and connected components are homogeneous from the forcing point of view.

that is, $s \in S$ iff $x_s^\alpha \in R$ for all α . So, for $\alpha \neq \beta$, $x_s^\alpha = x_s^\beta$ iff $s \in S$. Furthermore, if $s \in T_{\text{Hech}} \setminus S$, then $|s| \geq k+1$. List $\{t \in T \setminus S; t \upharpoonright (|t|-1) \in S\} = \{t_n; n \geq 1\}$. For $\alpha < \beta$ define

$$F(\{\alpha, \beta\}) = \begin{cases} \min\{n; & \text{either } t_n(|t_n|-1) \in \omega_1 \text{ and } x_{t_n}^\alpha(|t_n|-1) > x_{t_n}^\beta(|t_n|-1) \\ & \text{or } t_n(|t_n|-1) \in \omega_1^* \text{ and } x_{t_n}^\alpha(|t_n|-1) < x_{t_n}^\beta(|t_n|-1)\} \\ & \text{if such an } n \text{ exists,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that for each $n \geq 1$, every subset of λ_k homogeneous in color n must be finite. Using partition calculus as well as standard pruning arguments, we may therefore assume that for all $\alpha < \omega_1$, if $s \in S$ and $s^\wedge \langle \zeta \rangle \notin S$, then

- if ζ is positive, then $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) < x_{s^\wedge \langle \zeta \rangle}^\beta(|s|)$ for all $\alpha < \beta$, and all $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|)$ are larger than ω_1 , and if $\zeta < \xi$, $s^\wedge \langle \zeta \rangle, s^\wedge \langle \xi \rangle \notin S$, then
 - either for all α, β , we have $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) < x_{s^\wedge \langle \xi \rangle}^\beta(|s|)$ (this is the case when $\sup_\alpha x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) < \sup_\alpha x_{s^\wedge \langle \xi \rangle}^\alpha(|s|)$),
 - or for all $\alpha < \beta$, we have $x_{s^\wedge \langle \xi \rangle}^\alpha(|s|) < x_{s^\wedge \langle \zeta \rangle}^\beta(|s|)$ (this is the case when $\sup_\alpha x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) = \sup_\alpha x_{s^\wedge \langle \xi \rangle}^\alpha(|s|)$),
- if ζ is negative, then $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) > x_{s^\wedge \langle \zeta \rangle}^\beta(|s|)$ for all $\alpha < \beta$, and if $\zeta < \xi$, $s^\wedge \langle \zeta \rangle, s^\wedge \langle \xi \rangle \notin S$, then
 - either for all α, β , we have $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|) > x_{s^\wedge \langle \xi \rangle}^\beta(|s|)$,
 - or for all $\alpha < \beta$, we have $x_{s^\wedge \langle \xi \rangle}^\alpha(|s|) > x_{s^\wedge \langle \zeta \rangle}^\beta(|s|)$.

Define $x_s^\kappa \in L$ by recursion on the length of $s \in T$ as follows. If $s \in S$, then $x_s^\kappa = x_s^\alpha$ (a fortiori $|x_s^\kappa| = |x_s^\alpha| = |s|$). If $s \in S$ and $s^\wedge \langle \zeta \rangle \in T_{\text{Hech}} \setminus S$, let $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|)$ be the limit of the $x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|)$ (so it is either the sup or the inf, depending on whether ζ is positive or negative). Next find $\gamma_s < \lambda_{|s|+1}$, $\gamma_s > \omega_1$ and $\gamma_s^* \in S_{|s|+1}^{\theta_0}$, such that for all such s and ζ ,

- if $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|) = \sup_\alpha x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|)$, then for all $y \in B$ with $y \upharpoonright (|s|+1) = x_{s^\wedge \langle \zeta \rangle}^\kappa \upharpoonright (|s|+1)$, we have $y(|s|+1) > \gamma_s^*$,
- if $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|) = \inf_\alpha x_{s^\wedge \langle \zeta \rangle}^\alpha(|s|)$, then for all $y \in B$ with $y \upharpoonright (|s|+1) = x_{s^\wedge \langle \zeta \rangle}^\kappa \upharpoonright (|s|+1)$, we have $y(|s|+1) < \gamma_s$.

It is clear that we can find such γ_s 's because $\lambda_{|s|+1} > |B|$ is regular (since $|s| \geq k$). In the first case, let $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|+1) = \gamma_s^*$. In the second case, let $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|+1) = \gamma_s$. To complete the definition of $x_{s^\wedge \langle \zeta \rangle}^\kappa$, stipulate $|x_{s^\wedge \langle \zeta \rangle}^\kappa| = |x_{s^\wedge \langle \zeta \rangle}^\alpha| + 2 = |s| + 3$, and define

$$x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|+2) = \begin{cases} x_{s^\wedge \langle \zeta \rangle}^0(|s|) & \text{if } |s| > 0, \\ \zeta & \text{if } |s| = 0. \end{cases}$$

If $s \in S$ and $s^\wedge \langle \zeta \rangle \in T_{\text{mad}} \setminus S$, find $\gamma \in \lambda^* \cup \lambda$ such that $x_s^\kappa \upharpoonright \langle \gamma \rangle \in L_{\text{mad}}$ and for all $y \in B$ with $y \upharpoonright |s| = x_s^\kappa$, we have $y(|s|) \neq \gamma$. Such γ clearly exists because $\lambda > |B|$. Stipulate $|x_{s^\wedge \langle \zeta \rangle}^\kappa| = |x_s^\kappa| + 1 = |s| + 1$ and let $x_{s^\wedge \langle \zeta \rangle}^\kappa(|s|) = \gamma$. Finally, for the remaining $t \in T$, stipulate again $|x_t^\kappa| = |x_t^\alpha| + 2 = |t| + 2$, find $s \subset t$ with $s \in S$ maximal, and put $x_t^\kappa \upharpoonright (|s|+3) = x_{s^\wedge \langle t(|s|) \rangle}^\kappa$ and $x_t^\kappa(j+2) = x_t^0(j)$ for $j > |s|$.

Let $B^\kappa = \{x_s^\kappa; s \in T\}$. Notice that B^κ , although very tree-like, is not a tree like the B^α 's. We proceed to verify that $(B^\kappa, \mathcal{I} \upharpoonright B^\kappa)$ and $(B^\alpha, \mathcal{I} \upharpoonright B^\alpha)$, $\alpha < \omega_1$, are weakly isomorphic (clearly (a), (b) and (d) will fail but this is not relevant for us).

Fix α and define $\phi = \phi^{\alpha, \kappa}$ by $\phi(x_s^\alpha) = x_s^\kappa$. Without loss of generality, $\alpha = 0$. (c') and (e) are immediate by construction. So let us check that the trace of \mathcal{I} on B^0 is mapped to a cofinal subset of $\mathcal{I} \restriction B^\kappa$. First fix β and consider L_β . Notice that there is $\beta_0 \leq \beta$ such that $\phi(L_{\beta_0} \cap B^0) = L_{\beta_0} \cap B^\kappa = L_\beta \cap B^\kappa$. For any $s \in T$ with $x_s^0 \in L_\beta$ yet $x_s^\kappa \notin L_\beta$ one must have $x_s^\kappa(0) > \beta \geq x_s^0(0) \geq \beta_0$ and $x_s^\kappa(0)$ is an ω_1 -limit of the $x_s^\alpha(0)$'s. In particular, for all such s , $x_s^\kappa(0)$ must have the same value, say γ_0 . Also $x_s^\kappa(1) = \gamma_0^*$ and $x_s^\kappa(2) = s(0) < \omega_1$. This means, however, that $L_\beta \cap B^0$ is mapped to $(L_\beta \cup \text{cl}(J_x)) \cap B^\kappa$ via ϕ where $|x| = 3$, $x(0) = \gamma_0$, $x(1) = \gamma_0^*$ and $x(2) = \sup\{s(0) + 1; x_s^0(0) < \beta\}$. Next assume x is relevant and consider $\text{cl}(J_x)$. Assume $J_x \cap B^0 \neq \emptyset$. Then there is $s \in T$ such that $|s| = |x| - 1$ and $x_s^0 = x \restriction (|x| - 1)$. In case $s \in S$, we have $x_s^\kappa = x_s^0$ and $J_x \cap B^0$ is mapped to $J_x \cap B^\kappa$ via ϕ because $y \in R$ for any $y \in B^0$ with $|y| = |x|$, $y \restriction (|x| - 1) = x_s^0$ and $y(|x| - 1) \leq x(|x| - 1)$. In case $s \in T \setminus S$, let $j_0 < |s|$ be maximal with $s \restriction j_0 \in S$. Define y by $|y| = |x| + 2$, $y \restriction (|y| - 1) = x_s^\kappa$ and $y(|y| - 1) = x(|x| - 1)$ and note that $J_x \cap B^0$ gets mapped to $J_y \cap B^\kappa$ provided we can show y is relevant. In case $j_0 > 0$ there is nothing to show because whenever $x_s^0(j) > \omega_1$, $j \geq j_0$ even, then also $x_s^\kappa(j + 2) = x_s^0(j) > \omega_1$, and, if j_0 is even, we additionally have $x_s^\kappa(j_0) = \sup_\alpha x_s^\alpha(j_0) > \omega_1$ while, if j_0 is odd, we additionally have $x_s^\kappa(j_0 + 1) = \gamma_{s \restriction j_0} > \omega_1$. In case $j_0 = 0$ this is true because $x_s^\kappa(1) \in S_1^{\theta_0}$ and θ_0 is larger than all the θ for which $x_s^\kappa(j) \in S_j^\theta$ where $j > 1$ is odd.

Even though B^κ is not a tree, we can verify, as in the proof of Lemma 3.1, that, letting $C^\kappa := \text{cl}(B^\kappa)$, $C^\kappa \cap L_{\text{mad}} = \bigcup_{s \in T} \text{Comp}_{x_s^\kappa}$ and that $\psi = \psi^{\alpha, \kappa} : C^\alpha \rightarrow C^\kappa$, $\alpha < \omega_1$, which extends ϕ and maps $\text{Comp}_{x_s^\alpha}$ to $\text{Comp}_{x_s^\kappa}$ is a weak isomorphism.

By Lemma 3.2, $\mathbb{P} \restriction C^\alpha$ and $\mathbb{P} \restriction C^\kappa$ are isomorphic by a map $\chi = \chi^{\alpha, \kappa}$. χ sends $\mathbb{P} \restriction C^\alpha$ -names to $\mathbb{P} \restriction C^\kappa$ -names, and we define \dot{A}^κ to be the image of \dot{A}^α under χ .

By construction, it is then also immediate that whenever $\beta < \kappa$, we can find $\alpha < \omega_1$ such that $B^\kappa \cup B^\beta$ and $B^\alpha \cup B^\beta$ are weakly isomorphic via the mapping fixing nodes of B^β and sending the x_s^κ to the corresponding x_s^α , and such that this mapping identifies cofinal subsets of the traces of \mathcal{I} on the two sets.¹⁰ Again, this weak isomorphism canonically extends to a weak isomorphism of $C^\kappa \cup C^\beta$ and $C^\alpha \cup C^\beta$, which in turn means that $\mathbb{P} \restriction C^\kappa \cup C^\beta$ and $\mathbb{P} \restriction C^\alpha \cup C^\beta$ are isomorphic (Lemma 3.2) by a mapping sending the name \dot{A}^κ to \dot{A}^α . Since \dot{A}^α and \dot{A}^β are forced to be almost disjoint (by $\mathbb{P} \restriction C^\alpha \cup C^\beta$), so are \dot{A}^κ and \dot{A}^β (by the isomorphic $\mathbb{P} \restriction C^\kappa \cup C^\beta$). Since $\mathbb{P} \restriction C^\kappa \cup C^\beta$ embeds into $\mathbb{P} \restriction L$ (Lemma 1.1), this completes the proof of the non-maximality of \dot{A} and, by Corollary 2.2, of the Main Theorem. \square

REFERENCES

- [BJ] T. Bartoszyński and H. Judah, *Set Theory. On the structure of the real line*, A K Peters, Wellesley, MA, 1995. MR **96k**:03002
- [Bl] A. Blass, *Combinatorial cardinal characteristics of the continuum*, in: Handbook of Set Theory (A. Kanamori et al., eds.), to appear.
- [Br] J. Brendle, *Mad families and iteration theory*, in: Logic and Algebra (Y. Zhang, ed.), Contemp. Math. 302 (2002), Amer. Math. Soc., Providence, RI, 1-31.
- [H1] S. Hechler, *Short complete nested sequences in $\beta\mathbb{N} \setminus \mathbb{N}$ and small maximal almost-disjoint families*, General Topology and Appl. 2 (1972), 139-149. MR **46**:7028
- [H2] S. Hechler, *On the existence of certain cofinal subsets of ω^ω* , in: Axiomatic Set Theory (T. Jech, ed.), Proc. Sympos. Pure Math. 13 (1974), 155-173. MR **50**:12716
- [M] A. Miller, *Arnie Miller's problem list*, in: Set Theory of the Reals (H. Judah, ed.), Israel Math. Conf. Proc. 6 (1993), 645-654. MR **94m**:03073

¹⁰In fact, this is true for all but countably many $\alpha < \omega_1$.

- [S1] S. Shelah, *Covering of the null ideal may have countable cofinality*, Fund. Math. 166 (2000), 109-136. (publication number 592) MR **2001m**:03101
- [S2] S. Shelah, *Are \mathfrak{a} and \mathbb{D} your cup of tea?* Acta Math., to appear. (publication number 700)
- [vD] E. van Douwen, *The integers and topology*, in: Handbook of Set-theoretic Topology (K. Kunen and J. Vaughan, eds.), North-Holland, Amsterdam (1984), 111-167. MR **87f**:54008
- [V] J. E. Vaughan, *Small uncountable cardinals and topology*, in: Open Problems in Topology (J. van Mill and G. M. Reed, eds.), North-Holland (1990), 195-218.

THE GRADUATE SCHOOL OF SCIENCE AND TECHNOLOGY, KOBE UNIVERSITY, ROKKO-DAI 1-1,
NADA-KU, KOBE 657-8501, JAPAN

E-mail address: brendle@kurt.scitec.kobe-u.ac.jp

CYCLICITY OF CM ELLIPTIC CURVES MODULO p

ALINA CARMEN COJOCARU

ABSTRACT. Let E be an elliptic curve defined over \mathbb{Q} and with complex multiplication. For a prime p of good reduction, let \overline{E} be the reduction of E modulo p . We find the density of the primes $p \leq x$ for which $\overline{E}(\mathbb{F}_p)$ is a cyclic group. An asymptotic formula for these primes had been obtained conditionally by J.-P. Serre in 1976, and unconditionally by Ram Murty in 1979. The aim of this paper is to give a new simpler unconditional proof of this asymptotic formula and also to provide explicit error terms in the formula.

1. INTRODUCTION

Let E be an elliptic curve defined over \mathbb{Q} and of conductor N . By a famous result of Mordell, the set $E(\mathbb{Q})$ of \mathbb{Q} -rational points of E is a finitely generated abelian group. The study of the free part of $E(\mathbb{Q})$ is still one of the major problems in arithmetic geometry.

Now, for a prime p of good reduction for E (that is, $p \nmid N$), we denote by \overline{E} the reduction of E modulo p . This is an elliptic curve defined over \mathbb{F}_p , the finite field with p elements. Naturally, as in the rational case, one is interested in the study of the structure of the group $\overline{E}(\mathbb{F}_p)$ of \mathbb{F}_p -rational points of \overline{E} . From classical theory, $\overline{E}(\mathbb{F}_p)$ can be written as the product of two cyclic finite groups. Indeed, $\overline{E}(\mathbb{F}_p) \subseteq \overline{E}(\overline{\mathbb{F}_p})[k] \subseteq \mathbb{Z}/k\mathbb{Z} \oplus \mathbb{Z}/k\mathbb{Z}$, where $\overline{\mathbb{F}_p}$ denotes the algebraic closure of \mathbb{F}_p , k is a positive integer such that the order $\#\overline{E}(\mathbb{F}_p)$ of $\overline{E}(\mathbb{F}_p)$ divides k , and $\overline{E}(\overline{\mathbb{F}_p})[k]$ denotes the group of $\overline{\mathbb{F}_p}$ -rational points of \overline{E} annihilated by k . Early computations of Borosh, Moreno and Porta ([BMP]) showed that, in fact, for “many” primes p , the group $\overline{E}(\mathbb{F}_p)$ is cyclic. One expects this to be true for infinitely many primes p , as suggested by the elliptic curve analogue of Artin’s primitive root conjecture formulated by Lang and Trotter in 1977 (see [LT2]).

Our goal in this paper is to provide an asymptotic formula, with explicit error terms, for the function

$$f(x, \mathbb{Q}) := \#\{p \leq x : p \nmid N, \overline{E}(\mathbb{F}_p) \text{ cyclic}\},$$

in the case of an elliptic curve E defined over \mathbb{Q} and with complex multiplication.

In 1976 (see [Se1]), J. -P. Serre showed that C. Hooley’s *conditional* method of proving Artin’s conjecture on primitive roots (see [Ho, ch. 3]) can be adapted to estimate $f(x, \mathbb{Q})$. More precisely, let $\overline{\mathbb{Q}}$ denote the algebraic closure of \mathbb{Q} and

Received by the editors July 24, 2002 and, in revised form, December 4, 2002.

2000 *Mathematics Subject Classification.* Primary 11G05; Secondary 11N36, 11G15, 11R45.

Key words and phrases. Cyclicity of elliptic curves modulo p , complex multiplication, applications of sieve methods.

Research partially supported by an Ontario Graduate Scholarship.

let $\mathbb{Q}(E[k])$ denote the field obtained by adjoining to \mathbb{Q} the coordinates of the $\overline{\mathbb{Q}}$ -rational points of E annihilated by k . Then, under the Generalized Riemann Hypothesis (denoted GRH) for the Dedekind zeta functions of the division fields $\mathbb{Q}(E[k])$ of E , Serre proves that, as $x \rightarrow \infty$,

$$(1) \quad f(x, \mathbb{Q}) = f_E \operatorname{li} x + o\left(\frac{x}{\log x}\right),$$

where $\operatorname{li} x := \int_2^x \frac{1}{\log t} dt$ is the logarithmic integral and

$$f_E := \sum_{k=1}^{\infty} \frac{\mu(k)}{[\mathbb{Q}(E[k]) : \mathbb{Q}]},$$

with $\mu(\cdot)$ denoting the Möbius function.

We recall that for real-valued functions f and $g \neq 0$ we write $f(x) = o(g(x))$ to mean that $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$. Also, if g has positive values, we write $f(x) = O(g(x))$ or $f \ll g$ to mean that there exists a positive constant A such that $|f(x)| \leq Ag(x) \forall x$. If the constant A depends on some quantity B , then we may write $f(x) = O_B(g(x))$ or $f \ll_B g$. In this paper, whenever we write $f(x) = O(g(x))$ or $f \ll g$, we mean that the implied O -constants are absolute. If $f \ll g \ll f$, then we write $f \asymp g$.

In 1979¹ (see [Mu1, pp. 161-167]), Ram Murty removed GRH in formula (1) for elliptic curves with complex multiplication (denoted CM). His proof uses class field theoretical properties of the division fields of CM elliptic curves, as well as a number field version of the Bombieri-Vinogradov theorem (whose proof is based on the large sieve for number fields). In 2000 (see [acC1]), the author proved formula (1) for elliptic curves without complex multiplication (denoted non-CM) under the assumption of a quasi-GRH (more precisely, a zero-free region of real part $> 3/4$ for the Dedekind zeta functions of $\mathbb{Q}(E[k])$). For more history about $f(x, \mathbb{Q})$ in both the CM and non-CM cases we refer the reader to [acC1], [acC2] and [Mu3].

In this paper we give a new simpler *unconditional* proof for the asymptotic formula for $f(x, \mathbb{Q})$ in the complex multiplication case, and provide explicit error terms in this formula. We are proving the following:

Theorem 1.1. *Let E be a CM elliptic curve defined over \mathbb{Q} , of conductor N and with complex multiplication by the full ring of integers \mathcal{O}_K of an imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$, where D is a positive square-free integer. Then, as $x \rightarrow \infty$,*

$$(2) \quad f(x, \mathbb{Q}) = f_E \operatorname{li} x + O_N \left(\frac{x}{(\log x)(\log \log \log x)} \right),$$

or, more precisely,

$$(3) \quad f(x, \mathbb{Q}) = f_E \operatorname{li} x + O \left(\frac{x}{(\log x)(\log \log \frac{\log x}{N^2})} \cdot \frac{\log \log x}{\log \frac{\log x}{N^2}} \right),$$

where the O -constant in (2) depends on N and the one in (3) is absolute.

Corollary 1.2. *Let E be a CM elliptic curve defined over \mathbb{Q} , of conductor N and such that $\mathbb{Q}(E[2]) \neq \mathbb{Q}$. Then the smallest prime $p \nmid N$ for which $\overline{E}(\mathbb{F}_p)$ is cyclic has size $O(\exp(N^2))$. The implied O -constant is absolute.*

¹It was communicated to the author by Ram Murty that this result was obtained in 1979; however, it appeared in print only in 1983.

It is possible that the error terms in Theorem 1.1 can be improved, but this involves more sophisticated methods than the ones used in our paper. We relegate this to future research.

2. PRELIMINARIES

2.1. Notation. Given an elliptic curve E defined over \mathbb{Q} , p will always denote a prime of good reduction for E . We set $a_p := p + 1 - \# \overline{E}(\mathbb{F}_p)$ and say that p is of ordinary reduction if $a_p \neq 0$, and of supersingular reduction if $a_p = 0$. We denote by π_p and $\overline{\pi_p}$ the roots of the polynomial $X^2 - a_p X + p \in \mathbb{Z}[X]$.

If not otherwise stated, q will denote a rational prime and k a positive integer; $\pi(x)$ will denote the number of rational primes $\leq x$; $\#\mathcal{S}$ will denote the cardinality of a set \mathcal{S} ; $\text{Ker } \phi$ will denote the kernel of a morphism ϕ .

2.2. Algebraic preliminaries. The following preliminary lemmas are well known, but, for the sake of completeness, we include them here.

Lemma 2.1. *Let E be an elliptic curve defined over \mathbb{Q} and of conductor N . Let $E[k]$ be the group of k -division points of E . Then*

- (1) *the ramified primes of $\mathbb{Q}(E[k])/\mathbb{Q}$ are divisors of kN ;*
- (2) *assuming that E has complex multiplication and $k > 2$, we have*

$$\phi(k)^2 \ll [\mathbb{Q}(E[k]) : \mathbb{Q}] \ll k^2,$$

where $\phi(k)$ denotes the Euler function.

For proofs of this lemma the reader is referred to [Silv1, p. 179] and [Silv2, p. 135].

Lemma 2.2. *Let E be an elliptic curve defined over \mathbb{Q} and of conductor N . Using the notation introduced in Section 2.1 we have that, for a positive integer k and a prime $p \nmid k$ of good reduction for E , p splits completely in $\mathbb{Q}(E[k])/\mathbb{Q}$ if and only if $\frac{\pi_p - 1}{k}$ is an algebraic integer.*

Proof. We recall that π_p is the algebraic quadratic integer corresponding to the Frobenius endomorphism

$$\begin{aligned} \overline{E}(\overline{\mathbb{F}_p}) &\longrightarrow \overline{E}(\overline{\mathbb{F}_p}) \\ (x, y) &\mapsto (x^p, y^p), \end{aligned}$$

which we also denote by π_p .

Since $(p, kN) = 1$, we have that p is unramified in $\mathbb{Q}(E[k])/\mathbb{Q}$ (see part 1 of Lemma 2.1). By classical results in algebraic number theory, p splits completely in $\mathbb{Q}(E[k])/\mathbb{Q}$ if and only if $\pi_p|_{\overline{E}[k]} = 1$, where 1 denotes the identity map. This last condition is equivalent to saying that $\text{Ker}([k]) \subseteq \text{Ker}(\pi_p - 1)$ as maps $\overline{E}(\overline{\mathbb{F}_p}) \longrightarrow \overline{E}(\overline{\mathbb{F}_p})$, where $[k]$ is the multiplication by k map. Hence there exists an elliptic curve endomorphism $\phi : \overline{E}(\overline{\mathbb{F}_p}) \longrightarrow \overline{E}(\overline{\mathbb{F}_p})$ such that $\phi \circ [k] = \pi_p - 1$ (see [Silv1, Corollary 4.11, p. 77]). This is equivalent to saying that $\frac{\pi_p - 1}{k}$ is an algebraic integer. \square

Lemma 2.3. *Let E be a CM elliptic curve defined over \mathbb{Q} and with complex multiplication by an imaginary quadratic field K . Then, for every prime p of ordinary good reduction for E , we have $\mathbb{Q}(\pi_p) = K$.*

Proof. First we observe that

$$\mathbb{Q}(\pi_p) \subseteq \text{End}_{\mathbb{F}_p}(\overline{E}) \otimes_{\mathbb{Z}} \mathbb{Q} \subseteq \text{End}_{\overline{\mathbb{F}_p}}(\overline{E}) \otimes_{\mathbb{Z}} \mathbb{Q}.$$

Then we note that, since E has complex multiplication by K , we have an embedding $K \subseteq \text{End}_{\overline{\mathbb{F}_p}}(\overline{E}) \otimes_{\mathbb{Z}} \mathbb{Q}$, and, moreover, since p is a prime of ordinary reduction, we actually have $K = \text{End}_{\overline{\mathbb{F}_p}}(\overline{E}) \otimes_{\mathbb{Z}} \mathbb{Q}$. Thus $\mathbb{Q}(\pi_p) \subseteq K$ for any prime p of ordinary reduction for E . But K is a degree 2 extension of \mathbb{Q} , and so is $\mathbb{Q}(\pi_p)$. This gives us the desired equality. \square

Lemma 2.3 describes a feature of CM elliptic curves that will play a very important role in our unconditional estimates of $f(x, \mathbb{Q})$. It actually describes one of the main differences between CM and non-CM elliptic curves (see [LT1]).

2.3. Analytic preliminaries. The next preliminary lemma is an application of the sieve of Eratosthenes, which we recall below.

Theorem 2.4 (The sieve of Eratosthenes). *Let \mathcal{A} be a set of natural numbers $\leq x$, and let \mathcal{P} be a set of rational primes. To each prime $p \in \mathcal{P}$ we associate $\omega(p)$ distinguished residue classes modulo p . For any square-free integer d composed of primes of \mathcal{P} we set*

$$\mathcal{A}(d) := \{a \in \mathcal{A} : a \text{ belongs to at least one of the } \omega(p) \text{ residue classes modulo } p \text{ for all } p|d\},$$

and

$$\omega(d) := \prod_{p|d} \omega(p).$$

For a fixed real number z , we let $S(\mathcal{A}, \mathcal{P}, z)$ be the number of elements $a \in \mathcal{A}$ that do not belong to any of the distinguished residue classes modulo p for all $p \in \mathcal{P}, p \leq z$, and we set

$$W(z) := \prod_{\substack{p \in \mathcal{P} \\ p \leq z}} \left(1 - \frac{\omega(p)}{p}\right).$$

We assume that

- (1) *there exists a real number X such that, for all square-free integers d composed of primes of \mathcal{P} ,*

$$\#\mathcal{A}(d) = X \frac{\omega(d)}{d} + R_d$$

for some $R_d = O(\omega(d))$;

- (2) $\sum_{\substack{p \in \mathcal{P} \\ p \leq z}} \frac{\omega(p) \log p}{p} \leq c \log z + O(1)$ *for some positive constant c .*

Then

$$S(\mathcal{A}, \mathcal{P}, z) = XW(z) + O\left(x(\log z)^{c+1} \exp\left(-\frac{\log x}{\log z}\right)\right),$$

where the implied O -constant is absolute.

For a proof of this result, see [Mu4, p. 141].

Lemma 2.5. *Let $x \in \mathbb{R}$ and let D, k be fixed positive integers with $k < \sqrt{x} - 1$. Then*

$$\begin{aligned} S_k^1 &:= \#\{p \leq x : p = (\alpha k + 1)^2 + D\beta^2 k^2 \text{ for some } \alpha, \beta \in \mathbb{Z}\} \\ &= O\left(\left(\frac{\sqrt{x}}{k} + 1\right) \frac{\sqrt{x} \log \log x}{k\sqrt{D} \log \frac{\sqrt{x}-1}{k}}\right); \\ S_k^2 &:= \#\left\{p \leq x : p = \left(\frac{\alpha}{2}k + 1\right)^2 + D\frac{\beta^2}{4}k^2 \text{ for some } \alpha, \beta \in \mathbb{Z}\right\} \\ &= O\left(\left(\frac{\sqrt{x}}{k} + 1\right) \frac{\sqrt{x} \log \log x}{k\sqrt{D} \log \frac{\sqrt{x}-1}{k}}\right). \end{aligned}$$

The implied O -constants are absolute.

Proof. 1. Let us observe that the conditions $p \leq x$ and $p = (\alpha k + 1)^2 + D\beta^2 k^2$ for some $\alpha, \beta \in \mathbb{Z}$ imply

$$\begin{aligned} \alpha &\in \left[\frac{-1 - \sqrt{x}}{k}, \frac{-1 + \sqrt{x}}{k}\right] \cap \mathbb{Z}, \\ \beta &\in \left[-\frac{\sqrt{x}}{k\sqrt{D}}, \frac{\sqrt{x}}{k\sqrt{D}}\right] \cap \mathbb{Z}, \quad \beta \neq 0. \end{aligned}$$

Thus

$$(4) \quad S_k^1 \leq \sum_{\beta}' \#\left\{\alpha \in \left[\frac{-1 - \sqrt{x}}{k}, \frac{-1 + \sqrt{x}}{k}\right] \cap \mathbb{Z} : (\alpha k + 1)^2 + D\beta^2 k^2 \text{ a prime}\right\},$$

where the sum \sum_{β}' is over nonzero numbers $\beta \in \left[-\frac{\sqrt{x}}{k\sqrt{D}}, \frac{\sqrt{x}}{k\sqrt{D}}\right] \cap \mathbb{Z}$. We set

$$\begin{aligned} \mathcal{A} &:= \left\{\alpha \in \left[\frac{-1 - \sqrt{x}}{k}, \frac{-1 + \sqrt{x}}{k}\right] \cap \mathbb{Z}\right\}, \\ \mathcal{P} &:= \left\{p \text{ a rational prime} : (p, k) = 1, \left(\frac{-D}{p}\right) = 1\right\}, \end{aligned}$$

with $\left(\frac{\cdot}{p}\right)$ denoting the Legendre symbol modulo p . To each prime $p \in \mathcal{P}$ we associate the residue classes

$$(-1 \pm \beta k \mathcal{D})k^{-1} \pmod{p},$$

where \mathcal{D} is an integer such that $\mathcal{D}^2 \equiv -D \pmod{p}$ (let us observe that $(\alpha k + 1)^2 + D\beta^2 k^2 = p$ imposes the conditions $\left(\frac{-D}{p}\right) = 1$ and $(p, k) = 1$, and hence \mathcal{D} and $k^{-1} \pmod{p}$ are well defined).

For a fixed real number z we have

$$\begin{aligned} (5) \quad &\#\left\{\alpha \in \left[\frac{-1 - \sqrt{x}}{k}, \frac{-1 + \sqrt{x}}{k}\right] \cap \mathbb{Z} : (\alpha k + 1)^2 + D\beta^2 k^2 \text{ a prime}\right\} \\ &\leq S(\mathcal{A}, \mathcal{P}, z) + \pi(z) \\ &\leq S(\mathcal{A}, \mathcal{P}, z) + z, \end{aligned}$$

with $S(\mathcal{A}, \mathcal{P}, z)$ defined as in the sieve of Eratosthenes.

Now we want to verify that the hypotheses of Theorem 2.4 are satisfied. Elementary estimates give us

$$\#\mathcal{A}(d) := \#\{\alpha \in \mathcal{A} : (\alpha k + 1)^2 + D\beta^2 k^2 \equiv 0 \pmod{d}\} = 2 \left(\frac{2\sqrt{x}}{k} + 1 \right) \frac{1}{d} + O(1)$$

for all square-free integers d composed of primes of \mathcal{P} . Thus the first hypothesis of the sieve of Eratosthenes is satisfied with $\omega(d) = 2$ and $X = \frac{2\sqrt{x}}{k} + 1$. Using Mertens' theorem and recalling that $\left(\frac{-D}{p}\right) = 1$, hence that p splits completely in $\mathbb{Q}(\sqrt{-D})$, we obtain

$$\sum_{\substack{p \in \mathcal{P} \\ p \leq z}} \frac{\omega(p) \log p}{p} = 2 \sum_{\substack{p \in \mathcal{P} \\ p \leq z}} \frac{\log p}{p} = \log z + O(1).$$

Thus the second hypothesis of the sieve of Eratosthenes is satisfied with $c = 1$. Therefore,

$$S(\mathcal{A}, \mathcal{P}, z) = \left(\frac{2\sqrt{x}}{k} + 1 \right) W(z) + O \left(\frac{\sqrt{x} + 1}{k} (\log z)^2 \exp \left(-\frac{\log \frac{\sqrt{x}-1}{k}}{\log z} \right) \right),$$

where

$$W(z) = \prod_{\substack{p \in \mathcal{P} \\ p \leq z}} \left(1 - \frac{2}{p} \right) \leq \exp \left(-2 \sum_{\substack{p \in \mathcal{P} \\ p \leq z}} \frac{1}{p} \right) \ll \exp(-\log \log z) = \frac{1}{\log z},$$

by using the elementary inequality $1 + t \leq \exp(t)$ and, again, Mertens' theorem.

Let us choose z such that

$$\log z = \frac{\log \frac{\sqrt{x}-1}{k}}{3 \log \log x}.$$

Then

$$O \left(\frac{\sqrt{x} + 1}{k} (\log z)^2 \exp \left(-\frac{\log \frac{\sqrt{x}-1}{k}}{\log z} \right) \right) = O \left(\frac{\sqrt{x} + 1}{k} \frac{1}{\log x (\log \log x)^2} \right),$$

and so

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}, z) &= \left(\frac{2\sqrt{x}}{k} + 1 \right) O \left(\frac{\log \log x}{\log \frac{\sqrt{x}-1}{k}} \right) + O \left(\frac{\sqrt{x}}{k \log x (\log \log x)^2} \right) \\ &= \left(\frac{2\sqrt{x}}{k} + 1 \right) O \left(\frac{\log \log x}{\log \frac{\sqrt{x}-1}{k}} \right). \end{aligned}$$

From (5) we obtain

$$\begin{aligned} \# \left\{ \alpha \in \left[\frac{-1 - \sqrt{x}}{k}, \frac{-1 + \sqrt{x}}{k} \right] \cap \mathbb{Z} : (\alpha k + 1)^2 + D\beta^2 k^2 \text{ a prime} \right\} \\ = \left(\frac{2\sqrt{x}}{k} + 1 \right) O \left(\frac{\log \log x}{\log \frac{\sqrt{x}-1}{k}} \right), \end{aligned}$$

which, used in (4), completes the proof of the first part of the lemma.

2. Similar to the proof above.

□

We remark that for S_k^1 and S_k^2 of the above lemma we actually have elementary estimates that are weaker than the ones given by Lemma 2.5 only by a $\frac{\log \log x}{\log x}$ factor. The sieve has been invoked precisely for obtaining this saving.

Lemma 2.6. *Keeping the notation of Lemma 2.5, we have that, for any k and x ,*

$$S_k^i \ll \frac{\sqrt{x}}{k\sqrt{D}} \left(\frac{2\sqrt{x}}{k} + 1 \right),$$

where $1 \leq i \leq 2$.

Proof. We justify this estimate for $i = 1$. The case $i = 2$ is resolved similarly. We observe that the conditions $p \leq x$ and $p = (\alpha k + 1)^2 + D\beta^2 k^2$ for some $\alpha, \beta \in \mathbb{Z}$ give us $\frac{2\sqrt{x}}{k} + 1$ choices for α and $\frac{2\sqrt{x}}{k\sqrt{D}}$ choices for β . The lemma follows. \square

3. THE PROOF OF THE THEOREM AND COROLLARY

As explained in [Mu1, pp. 153-154], we have that $\overline{E}(\mathbb{F}_p)$ is cyclic if and only if p does not split completely in $\mathbb{Q}(E[q])$ for any prime $q \neq p$. Also, we have that if $p \leq x$ and p splits completely in $\mathbb{Q}(E[k])$ for some k , then $k^2 | (p + 1 - a_p)$, and so, using Hasse's bound $a_p \leq 2\sqrt{p}$, we obtain $k \leq 2\sqrt{x}$. Therefore, using the simple asymptotic sieve, we can write

$$f(x, \mathbb{Q}) = N(x, y) + O(M(x, y, 2\sqrt{x})),$$

where

$$N(x, y) := \#\{p \leq x : p \text{ does not split completely in any } \mathbb{Q}(E[q])/\mathbb{Q}, q \leq y\},$$

$$M(x, y, 2\sqrt{x}) := \#\{p \leq x : p \text{ splits completely in some } \mathbb{Q}(E[q])/\mathbb{Q} \\ \text{with } y \leq q \leq 2\sqrt{x}\},$$

and where y is a real number to be chosen later. In order to estimate $f(x, \mathbb{Q})$ we need to estimate each of $N(x, y)$ and $M(x, y, 2\sqrt{x})$ and to choose the parameter y appropriately.

3.1. Estimate for $N(x, y)$. By the inclusion-exclusion principle we have

$$N(x, y) = \sum_k' \mu(k) \pi_1(x, \mathbb{Q}(E[k])/\mathbb{Q}),$$

where the sum is over all square-free positive integers $k \leq 2\sqrt{x}$ whose prime divisors are $\leq y$, and where

$$\pi_1(x, \mathbb{Q}(E[k])/\mathbb{Q}) := \#\{p \leq x : p \text{ splits completely in } \mathbb{Q}(E[k])/\mathbb{Q}\}.$$

We estimate this sum by using the unconditional effective version of the Chebotarev density theorem as stated in [Mu2, p. 243] or [acC1, p. 337]. To do so, let us recall from [Se2, p. 130] that if L/\mathbb{Q} is a finite normal field extension that is ramified only at the primes p_1, p_2, \dots, p_m , then

$$\frac{1}{[L : \mathbb{Q}]} \log |\text{disc}(L/\mathbb{Q})| \leq \log[L : \mathbb{Q}] + \sum_{j=1}^m \log p_j,$$

where $[L : \mathbb{Q}]$ and $\text{disc}(L/\mathbb{Q})$ denote the degree and the discriminant, respectively, of L/\mathbb{Q} . We apply this result, together with Lemma 2.1, to the fields $\mathbb{Q}(E[k])$, whose degree and discriminant over \mathbb{Q} we denote by $n(k)$ and d_k , respectively. We get

$$n(k)|d_k|^{\frac{2}{n(k)}} \ll k^8 N^2$$

and

$$n(k) (\log |d_k|)^2 \ll k^6 (\log (k^2 N))^2,$$

and so the maximum of the two quantities above is $\ll k^8 N^2$. In order to apply the unconditional effective Chebotarev density theorem mentioned before we need to have $k^8 N^2 \ll \log x$. Since $k \leq \exp(2y)$, it is enough to choose

$$(6) \qquad y = \frac{1}{8}(\log \log x - 2 \log N).$$

Then, by the unconditional effective Chebotarev density theorem, we obtain

$$N(x, y) = \left(\sum_k' \frac{\mu(k)}{n(k)} \right) \text{li } x + O \left(\sum_k' x \exp \left(-A \sqrt{\frac{\log x}{n(k)}} \right) \right)$$

for some effective positive constant A . To handle the error term we use that $n(k) \ll k^2$ and that there are at most 2^y square-free numbers composed of primes $\leq y$. Then

$$(7) \qquad N(x, y) = \left(\sum_k' \frac{\mu(k)}{n(k)} \right) \text{li } x + O \left(\frac{x}{N^{1/4}(\log x)^B} \right)$$

for any positive constant B .

3.2. Estimate for $M(x, y, 2\sqrt{x})$. For real numbers ξ_1, ξ_2 , we denote by

$$M^o(x, \xi_1, \xi_2)$$

the number of primes $p \leq x$ such that p has ordinary reduction and splits completely in some $\mathbb{Q}(E[q])$ with $\xi_1 \leq q \leq \xi_2$, and by

$$M^s(x, \xi_1, \xi_2)$$

the number of primes $p \leq x$ such that p has supersingular reduction and splits completely in some $\mathbb{Q}(E[q])$ with $\xi_1 \leq q \leq \xi_2$. We write

$$(8) \qquad M(x, y, 2\sqrt{x}) = M^o(x, y, 2\sqrt{x}) + M^s(x, y, 2\sqrt{x})$$

and estimate each of the two terms. For the first one we observe that

$$(9) \qquad M^o(x, y, 2\sqrt{x}) \leq \sum_{y < q \leq 2\sqrt{x}} \pi_1^o(x, \mathbb{Q}(E[q])/\mathbb{Q}),$$

where

$$\pi_1^o(x, \mathbb{Q}(E[q])/\mathbb{Q}) := \#\{p \leq x : a_p \neq 0 \text{ and } p \text{ splits completely in } \mathbb{Q}(E[q])/\mathbb{Q}\}.$$

By Lemmas 2.2 and 2.3 we obtain

$$\pi_1^o(x, \mathbb{Q}(E[q])/\mathbb{Q}) \leq \#\left\{p \leq x : \frac{\pi_p - 1}{q} \in \mathcal{O}_K\right\}.$$

Since the norm of π_p in K/\mathbb{Q} is p , we get

$$\#\left\{p \leq x : \frac{\pi_p - 1}{q} \in \mathcal{O}_K\right\} \leq S_q,$$

where S_q is S_q^1 if $-D \equiv 2, 3 \pmod{4}$, and S_q^2 if $-D \equiv 1 \pmod{4}$, with S_q^1, S_q^2 as in Lemma 2.5.

Let us fix a real number $u < \sqrt{x} - 1$. Using the elementary estimate for S_q given in Lemma 2.6, we obtain

$$\begin{aligned}
 \sum_{u < q \leq 2\sqrt{x}} \pi_1^o(x, \mathbb{Q}(E[q])/\mathbb{Q}) &\leq \sum_{u < q \leq 2\sqrt{x}} S_q \\
 &\ll \sum_{u < q \leq 2\sqrt{x}} \frac{\sqrt{x}}{q\sqrt{D}} \left(\frac{2\sqrt{x}}{q} + 1 \right) \\
 &= \frac{2x}{\sqrt{D}} \sum_{u < q \leq 2\sqrt{x}} \frac{1}{q^2} + \frac{\sqrt{x}}{\sqrt{D}} \sum_{u < q \leq 2\sqrt{x}} \frac{1}{q} \\
 (10) \qquad &\ll \frac{x}{\sqrt{D}u \log u} + \frac{\sqrt{x} \log \log x}{\sqrt{D}}.
 \end{aligned}$$

On the other hand, using the estimates for S_q given in Lemma 2.5, we obtain

$$\begin{aligned}
 \sum_{y < q \leq u} \pi_1^o(x, \mathbb{Q}(E[q])/\mathbb{Q}) &\leq \sum_{y < q \leq u} S_q \\
 &\ll \sum_{y \leq q \leq u} \left(\frac{x}{q^2\sqrt{D}} + \frac{\sqrt{x}}{q\sqrt{D}} \right) \frac{\log \log x}{\log \frac{\sqrt{x}-1}{q}} \\
 &\ll \frac{x \log \log x}{\sqrt{D} \log \frac{\sqrt{x}-1}{u}} \sum_{y < q \leq u} \frac{1}{q^2} + \frac{\sqrt{x} \log \log x}{\sqrt{D} \log \frac{\sqrt{x}-1}{u}} \sum_{y < q \leq u} \frac{1}{q} \\
 (11) \qquad &\ll \frac{x \log \log x}{\sqrt{D}(\log \frac{\sqrt{x}-1}{u})y \log y} + \frac{\sqrt{x}(\log \log x)(\log \log u)}{\sqrt{D} \log \frac{\sqrt{x}-1}{u}}.
 \end{aligned}$$

We choose

$$u = \log x$$

and recall that $y = \frac{1}{8}(\log \log x - 2 \log N)$ (see (6)) and that D is bounded, since E has CM. Then, from (9), (10) and (11) we get

$$(12) \qquad M^o(x, y, 2\sqrt{x}) = O \left(\frac{x \log \log x}{(\log x)(\log \frac{\log x}{N^2})(\log \log \frac{\log x}{N^2})} \right).$$

For the second term in (8) we have

$$(13) \qquad M^s(x, y, 2\sqrt{x}) \leq \sum_{y < q \leq 2\sqrt{x}} \pi_1^s(x, \mathbb{Q}(E[q])/\mathbb{Q}),$$

where

$$\pi_1^s(x, \mathbb{Q}(E[q])/\mathbb{Q}) := \#\{p \leq x : a_p = 0 \text{ and } p \text{ splits completely in } \mathbb{Q}(E[q])/\mathbb{Q}\}.$$

We observe that if p is a prime of supersingular reduction that splits completely in some $\mathbb{Q}(E[q])/\mathbb{Q}$, then $q = 2$. Indeed, for such primes p and q we have, on the one hand, that $q^2|(p+1-a_p) = (p+1)$, and, on the other hand, that $q|(p-1)$; thus $q|2$. Now we note that in the sum of (13) we run over $q > y$; thus, by our choice of y (see (6)), $q \neq 2$. This implies that

$$(14) \qquad M^s(x, y, 2\sqrt{x}) = 0.$$

3.3. The final formula. Putting together (7), (8), (12) and (14) we get

$$\begin{aligned} f(x, \mathbb{Q}) &= \left(\sum_k' \frac{\mu(k)}{n(k)} \right) \text{li } x + O\left(\frac{x}{N^{1/4}(\log x)^B} \right) \\ &\quad + O\left(\frac{x}{(\log x)(\log \log x)} \right) \\ &\quad + O\left(\frac{x}{(\log x)(\log \log \frac{\log x}{N^2})} \cdot \frac{\log \log x}{\log \frac{\log x}{N^2}} \right), \end{aligned}$$

where the implied O-constants are absolute. It remains to analyze $\left(\sum_k' \frac{\mu(k)}{n(k)} \right) \text{li } x$.

We write

$$\sum_k' \frac{\mu(k)}{n(k)} = \sum_k \frac{\mu(k)}{n(k)} - \sum_k'' \frac{\mu(k)}{n(k)},$$

where \sum_k'' means that the sum is over those positive square-free integers k for which there exists a prime divisor $q > y$. Using part 2 of Lemma 2.1 we get that

$$\begin{aligned} \sum_k'' \frac{\mu(k)}{n(k)} \text{li } x &\ll \frac{x}{\log x} \sum_{q>y} \sum_{t=1}^{\infty} \frac{1}{q^2 t^{3/2}} \\ &\ll \frac{x}{(\log x)y \log y} \\ &= O\left(\frac{x}{(\log x)(\log \frac{\log x}{N^2})(\log \log \frac{\log x}{N^2})} \right). \end{aligned}$$

Thus

$$(15) \quad f(x, \mathbb{Q}) = f_E \text{li } x + O\left(\frac{x}{(\log x)(\log \log \frac{\log x}{N^2})} \cdot \frac{\log \log x}{\log \frac{\log x}{N^2}} \right).$$

This completes the proof of Theorem 1.1.

3.4. The proof of Corollary 1.2. First, let us recall that it was pointed out by Serre (see [Mu3, p. 327]) that the density f_E is positive if and only if $\mathbb{Q}(E[2]) \neq \mathbb{Q}$. Now, we note that a necessary condition for formula (15) to hold is that $x \geq \exp(N^2)$. Then, if $x \asymp \exp(N^2)$, the main term of (15) will be bigger than the error term. This proves the assertion of the corollary.

4. CONCLUDING REMARKS

As mentioned in the proof of Corollary 1.2, the density f_E is positive if and only if $\mathbb{Q}(E[2]) \neq \mathbb{Q}$. For the sake of clarity, we explain this in what follows in the case of a CM elliptic curve E defined over \mathbb{Q} . Naturally, in order to have $f_E \neq 0$ we need to assume $\mathbb{Q}(E[2]) \neq \mathbb{Q}$, for otherwise the torsion part of $E(\mathbb{Q})$ contains the Klein four group and so $\overline{E}(\mathbb{F}_p)$ cannot be cyclic. The condition is also sufficient. To see this, let us first note that if $\mathbb{Q}(E[2]) \neq \mathbb{Q}$, then $[\mathbb{Q}(E[2]) : \mathbb{Q}]$ is 2, 3 or 6. We let K_2 be the unique abelian subextension contained in $\mathbb{Q}(E[2])$. Also, we let K be the CM field of E . We recall that $K(E[q]) = \mathbb{Q}(E[q])$ for any prime $q \geq 3$ (see [Mu1, p. 165, Lemma 6]), and we observe that since K is a quadratic field and K_2 is a cubic or a quadratic field, we have either $K_2 \cap K = \mathbb{Q}$ or $K_2 = K$. If $K_2 \cap K = \mathbb{Q}$,

then using that $K_2 \subseteq \mathbb{Q}(E[2])$ and $K \subseteq \mathbb{Q}(E[q])$ for any $q \geq 3$, we deduce that the density of the primes p that do not split completely in any of the fields $\mathbb{Q}(E[q])$ is greater than or equal to the density of the primes p that do not split completely in K_2 and K . In other words,

$$f_E \geq \left(1 - \frac{1}{[K_2 : \mathbb{Q}]}\right) \left(1 - \frac{1}{[K : \mathbb{Q}]}\right) \geq \frac{1}{4}.$$

If $K_2 = K$, then $K \subseteq \mathbb{Q}(E[q])$ for any prime q , and so the density of the primes p that do not split completely in any of the fields $\mathbb{Q}(E[q])$ is greater than or equal to the density of the primes p that do not split completely in K . In other words,

$$f_E \geq \left(1 - \frac{1}{[K : \mathbb{Q}]}\right) \geq \frac{1}{2}.$$

This completes the proof of the positivity of f_E .

The main significance of our unconditional proof of the asymptotic formula for $f(x, \mathbb{Q})$ in the case of a CM elliptic curve lies in the simplicity of the tools that are used. Ram Murty's initial proof avoided the GRH by using a difficult application of the large sieve for number fields, namely a Bombieri-Vinogradov type result for number fields. In our new proof we use instead an application of the sieve of Eratosthenes, one of the simplest sieves in number theory. We point out that this application of the sieve of Eratosthenes (Lemma 2.5) could be viewed as a Brun-Titchmarsh theorem for quadratic number fields, since it gives nontrivial upper bounds for the number of (principal) prime ideals whose generator satisfies congruence conditions. A result of this kind had been obtained in [Sch], but as an application of the large sieve for number fields, and could have been used in our treatment of $M(x, y, 2\sqrt{x})$.

Another significance of our new proof is that it provides explicit error terms, with absolute O -constants. As noted in Corollary 1.2, we can then deduce an unconditional upper bound for the smallest prime p for which $\overline{E}(\mathbb{F}_p)$ is cyclic. Considerable improvements of this bound, under GRH, will be discussed in an upcoming paper.

Naturally, one can ask if our ideas can be explored further and used in other related situations. For example, one could consider the question of determining the number of prime ideals for which the reduction of a CM elliptic curve defined over a number field gives a cyclic group. It seems that our tools can be used in this situation. Another question is that of using the ideas of this paper in the case of a non-CM elliptic curve. At present, no unconditional proof for the asymptotic formula for $f(x, \mathbb{Q})$ is known in this situation, but, as mentioned in Section 1, only a proof based on a quasi-GRH assumption (see [acC1]). If we assume a variation of a conjecture of Lang and Trotter on the number of distinct fields $\mathbb{Q}(\pi_p)$ obtained when p runs over primes of ordinary reduction for a non-CM elliptic curve (see [LT1]), then it turns out that we can follow the current CM approach even in the non-CM case. The dependence $\frac{1}{\sqrt{D}}$ on the discriminant D of the estimates provided by Lemma 2.5 will be more advantageous than the dependence on D provided by Schaal's result mentioned above. This is, again, an asset of our new proof. Yet another related question is that of determining an asymptotic formula for the number of primes p for which the order of $\overline{E}(\mathbb{F}_p)$ is square-free. The ideas of our paper can be successfully used to answer this question if E is a CM elliptic curve. The details of our last two claims will be given in different upcoming papers.

ACKNOWLEDGEMENTS

The results of this paper are part of my doctoral thesis [acC2]. I express my deepest gratitude to my supervisor, Professor M. Ram Murty, for all his help and support. I am also grateful to Professor Ernst Kani for useful discussions on the algebraic preliminaries of the paper.

REFERENCES

- [acC1] A. C. Cojocaru, "On the cyclicity of the group of \mathbb{F}_p -rational points of non-CM elliptic curves", *Journal of Number Theory*, vol. 96, no. 2, October 2002, pp. 335-350.
- [acC2] A. C. Cojocaru, "Cyclicity of elliptic curves modulo p ", Ph.D. thesis, Queen's University, Kingston, Canada, 2002.
- [BMP] I. Borosh, C. J. Moreno, and H. Porta, "Elliptic curves over finite fields II", *Mathematics of Computation*, vol. 29, July 1975, pp. 951-964. MR **53**:8067
- [Ho] C. Hooley, "Applications of sieve methods to the theory of numbers", Cambridge University Press, 1976. MR **53**:7976
- [LT1] S. Lang and H. Trotter, "Frobenius distributions in GL_2 -extensions", *Lecture Notes in Mathematics* 504, Springer-Verlag, 1976. MR **58**:27900
- [LT2] S. Lang and H. Trotter, "Primitive points on elliptic curves", *Bulletin of the American Mathematical Society*, vol. 83, no. 2, March 1977, pp. 289-292. MR **55**:308
- [Mu1] M. Ram Murty, "On Artin's conjecture", *Journal of Number Theory*, vol. 16, no. 2, April 1983, pp. 147-168. MR **86f**:11087
- [Mu2] M. Ram Murty, "An analogue of Artin's conjecture for abelian extensions", *Journal of Number Theory*, vol. 18, no. 3, June 1984, pp. 241-248. MR **85j**:11161
- [Mu3] M. Ram Murty, "Artin's conjecture and elliptic analogues", *Sieve Methods, Exponential Sums and their Applications in Number Theory* (eds. G. R. H. Greaves, G. Harman, M. N. Huxley), Cambridge University Press, 1996, pp. 326-344. MR **2000a**:11098
- [Mu4] M. Ram Murty, "Problems in analytic number theory", *Graduate Texts in Mathematics* 206, Springer-Verlag, 2001. MR **2001k**:11002
- [Sch] W. Schaal, "On the large sieve method in algebraic number fields", *Journal of Number Theory* 2, 1970, pp. 249-270. MR **42**:7626
- [Se1] J. -P. Serre, "Résumé des cours de 1977-1978", *Annuaire du Collège de France* 1978, pp. 67-70.
- [Se2] J. -P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Etudes Sci. Publ. Math.*, no. 54, 1981, pp. 123-201. MR **83k**:12011
- [Silv1] J. H. Silverman, "The arithmetic of elliptic curves", *Graduate Texts in Mathematics* 106, Springer-Verlag, New York, 1986. MR **87g**:11070
- [Silv2] J. H. Silverman, "Advanced topics in the arithmetic of elliptic curves", *Graduate Texts in Mathematics* 151, Springer-Verlag, New York, 1994. MR **96b**:11074

DEPARTMENT OF MATHEMATICS AND STATISTICS, QUEEN'S UNIVERSITY, KINGSTON, ONTARIO, CANADA, K7L 3N6

E-mail address: `alina@mast.queensu.ca`

Current address: The Fields Institute for Research in Mathematical Sciences, 222 College Street, Toronto, Ontario, M5T 3J1, Canada

E-mail address: `alina@fields.utoronto.ca`

TAYLOR EXPANSION OF AN EISENSTEIN SERIES

TONGHAI YANG

ABSTRACT. In this paper, we give an explicit formula for the first two terms of the Taylor expansion of a classical Eisenstein series of weight $2k + 1$ for $\Gamma_0(q)$. Both the first term and the second term have interesting arithmetic interpretations. We apply the result to compute the central derivative of some Hecke L -functions.

0. INTRODUCTION

Consider the classical Eisenstein series

$$\sum_{\gamma \in \Gamma_\infty \backslash \mathrm{SL}_2(\mathbb{Z})} \mathrm{Im}(\gamma\tau)^s,$$

which has a simple pole at $s = 1$. The well-known Kronecker limit formula gives a closed formula for the next term (the constant term) in terms of the Dedekind η -function and has a lot of applications in number theory. It seems natural and worthwhile to study the same question for more general Eisenstein series. For example, consider the Eisenstein series

$$(0.1) \quad E(\tau, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma_0(q)} \epsilon(d)(c\tau + d)^{-2k-1} \mathrm{Im}(\gamma\tau)^{\frac{s}{2}-k}.$$

Here $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $-q$ is a fundamental discriminant of an imaginary quadratic field, and $\epsilon = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This Eisenstein series was used in the celebrated work of Gross and Zagier ([GZ, Chapter IV]) to compute the central derivative of cuspidal modular forms of weight $2k + 2$. The Eisenstein series is holomorphic (as a function of s) at the symmetric center $s = 0$ with the leading term (constant term) given by a theta series via the Siegel-Weil formula. The analogue of the Kronecker limit formula would be a closed formula for the central derivative at $s = 0$ —the main object of this paper. This would give a direct proof of [GZ, Proposition 4.5]. Another application is to give a closed formula for the central derivative of a family of Hecke L -series associated to CM abelian varieties, which is very important in the arithmetic of CM abelian varieties in view of the Birch and Swinnerton-Dyer conjecture. This application will be given in section 4. We will also prove a transformation equation for the tangent line of the Eisenstein series at the center, which should be of independent interest.

Received by the editors September 9, 2002.

2000 *Mathematics Subject Classification.* Primary 11G05, 11M20, 14H52.

Key words and phrases. Kronecker formula, central derivative, elliptic curves, Eisenstein series. Partially supported by an AMS Centennial fellowship and NSF grant DMS-0070476.

To make the exposition simple, we assume that $q > 3$ is a prime congruent to 3 modulo 4. Let $\mathbf{k} = \mathbb{Q}(\sqrt{-q})$.

Set

$$(0.2) \quad \Lambda(s, \epsilon) = \pi^{-\frac{s+1}{2}} \Gamma\left(\frac{s+1}{2}\right) L(s, \epsilon)$$

and

$$(0.3) \quad E^*(\tau, s) = q^{\frac{s+1}{2}} \Lambda(s+1, \epsilon) E(\tau, s).$$

It is well known that $E^*(\tau, s)$ is holomorphic.

As in [GZ, Propositions 4.4 and 3.3], we define

$$(0.4) \quad p_k(t) = \sum_{m=0}^k \binom{k}{m} \frac{(-t)^m}{m!}$$

and

$$(0.5) \quad q_k(t) = \int_1^\infty e^{-tu} (u-1)^k u^{-k-1} du, \quad t > 0.$$

We remark that $p_k(-t)$ and $q_k(t)$ are two “basic” solutions of the differential equations

$$(0.6) \quad tC''(t) + (1+t)C'(t) - kC(t) = 0.$$

Finally, let $\rho(n)$ be given by

$$(0.7) \quad \zeta_{\mathbf{k}}(s) = \sum \rho(n) n^{-s}.$$

Theorem 0.1. *Let the notation be as above, and let h be the ideal class number of \mathbf{k} . Write $\tau = u + iv$. Then*

$$E^*(\tau, 0) = v^{-k} (h + 2 \sum_{n>0} \rho(n) p_k(4\pi n v) e(n\tau))$$

and

$$\begin{aligned} & E^{*'}(\tau, 0) + \frac{1}{4} \sum_{j=1}^k \frac{1}{j} E^*(\tau, 0) \\ &= \frac{1}{2} v^{-k} \left[a_0(v) - 2 \sum_{n>0} a_n p_k(4\pi n v) e(n\tau) - 2 \sum_{n<0} \rho(-n) q_k(-4\pi n v) e(n\tau) \right]. \end{aligned}$$

Here

$$a_0(v) = h \left(\log(qv) + 2 \frac{\Lambda'(1, \epsilon)}{\Lambda(1, \epsilon)} + \sum_{j=1}^k \frac{1}{j} \right)$$

and

$$a_n = (\text{ord}_q n + 1) \rho(n) \log q + \sum_{\left(\frac{n}{q}\right)=-1} (\text{ord}_p n + 1) \rho(n/p) \log p.$$

The formulas should be compared to those for $\tilde{\Phi}$ in [GZ, Propositions 4.4 and 4.5]. In fact, multiplying our formulas by the theta function in their paper and taking the trace would yield their formulas for $\tilde{\Phi}$. The method used here seems to be more suitable for generalization. The proof is based on the observation that the Eisenstein series (0.1) can be split into two Eisenstein series. One of them is coherent, and it is easy to compute its value. It contributes little to the central

derivative. The other one is incoherent, contributes nothing to the value, and its central derivative can be computed by the method of [KRY], where we dealt with the case $k = 0$. This consists of sections 1 and 2.

In section 3, we study how the value and derivative behave under the Fricke involution $\tau \mapsto -1/q\tau$ and obtain the following functional equation. One interesting point about the equation is that it basically follows from the definition of automorphic forms (see (3.2)).

Theorem 0.2. *The modular forms $E^*(\tau, 0)$ and $E^{*'}(\tau, 0)$ satisfy the following functional equation:*

$$\begin{pmatrix} E^*(-\frac{1}{q\tau}, 0) \\ E^{*'}(-\frac{1}{q\tau}, 0) \end{pmatrix} = i(\sqrt{q}\tau)^{2k+1} \begin{pmatrix} -1 & 0 \\ \sum_{j=1}^k \frac{1}{j} + \frac{1}{2} \log q & 1 \end{pmatrix} \begin{pmatrix} E^*(\tau, 0) \\ E^{*'}(\tau, 0) \end{pmatrix}.$$

Finally, let μ be a canonical Hecke character of weight 1 of \mathbf{k} (see section 4 for the definition). It is associated to the CM elliptic curve $A(q)$ studied by Gross ([Gro]). When $q \equiv 3 \pmod{8}$, S. Miller and the author proved recently that the central derivative $L'(1, \mu) \neq 0$ ([MY]). Since the central derivative encodes very important information in the arithmetic of $A(q)$, it is important to find a good formula for the central derivative. Standard calculation shows that the L -series $L(s, \mu)$ is $E(\tau, 2s)$ evaluated at a CM cycle. So Theorem 0.1 gives an explicit formula for the central derivative $L'(1, \mu)$ (Corollary 4.2).

1. COHERENT AND INCOHERENT EISENSTEIN SERIES

Let $G = \mathrm{SL}_2$ over \mathbb{Q} , and let $B = TN$ be the standard Borel subgroup, where T is the standard maximal split torus of B and N is the unipotent radical of B . Their rational points are given by

$$T(\mathbb{Q}) = \{m(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a \in \mathbb{Q}^*\}$$

and

$$N(\mathbb{Q}) = \{n(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{Q}\}.$$

Consider the global induced representation

$$I(s, \epsilon) = \mathrm{Ind}_{B(\mathbb{A})}^{G(\mathbb{A})} \epsilon | \cdot |_s$$

of $G(\mathbb{A})$, where \mathbb{A} is the ring of adèles of \mathbb{Q} . By definition a section $\Phi(s) \in I(s, \epsilon)$ satisfies

$$(1.1) \quad \Phi(n(b)m(a)g, s) = \epsilon(a)|a|^{s+1}\Phi(g, s)$$

for $a \in \mathbb{A}^*$ and $b \in \mathbb{A}$. Let $K = \mathrm{SL}_2(\hat{\mathbb{Z}})$ and let $K_\infty = \mathrm{SO}(2)(\mathbb{R})$. Associated to a standard section Φ , which means that its restriction on KK_∞ is independent of s , one defines the Eisenstein series

$$(1.2) \quad E(g, s, \Phi) = \sum_{\gamma \in B(\mathbb{Q}) \backslash G(\mathbb{Q})} \Phi(\gamma g, s).$$

It is absolutely convergent for $\mathrm{Re} s > 1$ and has a meromorphic continuation to the whole complex s -plane. We consider three standard sections Φ^0, Φ^\pm in this paper. For every prime $p \nmid q\infty$, let $\Phi_p \in I(s, \epsilon_p)$ be the unique spherical section

such that $\Phi_p(x) = 1$ for every $x \in K_p = \mathrm{SL}_2(\mathbb{Z}_p)$. Let $\Phi_\infty \in I(s, \epsilon_\infty)$ be the unique section of weight $2k + 1$ in the sense that

$$(1.3) \quad \Phi_\infty(gk_\theta, s) = \Phi_\infty(g, s)e^{i(2k+1)\theta}$$

for every $k_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \in K_\infty$. For $p = q$, let

$$J_q = \left\{ \begin{pmatrix} a & b \\ cq & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}_q) : a, b, c, d \in \mathbb{Z}_q \right\}$$

be the Iwahori subgroup of K_q . Then ϵ_q defines a character of J_q via

$$(1.4) \quad \epsilon_q\left(\begin{pmatrix} a & b \\ cq & d \end{pmatrix}\right) = \epsilon_q(d).$$

As described in [KRY, section 2], the subspace of $I(s, \epsilon_q)$ consisting of ϵ_q eigenvectors of J_q is two-dimensional and is spanned by the cell functions of Φ_q^i , determined by

$$(1.5) \quad \Phi_q^i(w_j, s) = \delta_{ij}, \quad \text{where } w_0 = 1 \text{ and } w_1 = w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We denote this subspace by $W(J_q, \epsilon_q, s)$. A better basis for this subspace turns out to be given by

$$(1.6) \quad \Phi_q^\pm = \Phi_q^0 \pm \frac{1}{\sqrt{-q}} \Phi_q^1,$$

which are “eigenfunctions” of some intertwining operator (see Lemma 2.2). Set

$$(1.7) \quad \Phi^0 = \Phi_q^0 \prod_{p \neq q} \Phi_p \quad \text{and} \quad \Phi^\pm = \Phi_q^\pm \prod_{p \neq q} \Phi_p.$$

Clearly, $\Phi^0 = \frac{1}{2}(\Phi^+ + \Phi^-)$. For $\tau = u + iv$ with $v > 0$, let

$$(1.8) \quad g_\tau = n(u)m(\sqrt{v}).$$

Then standard computation gives

Proposition 1.1. *Let the notation be as above. Then*

$$\begin{aligned} E^*(\tau, s) &= v^{-k-\frac{1}{2}} E^*(g_\tau, s, \Phi^0) \\ &= \frac{1}{2} v^{-k-\frac{1}{2}} (E^*(g_\tau, s, \Phi^+) + E^*(g_\tau, s, \Phi^-)). \end{aligned}$$

Here

$$E^*(g, s, \Phi) = q^{\frac{s+1}{2}} \Lambda(s+1, \epsilon) E(g, s, \Phi)$$

is the completion of the Eisenstein series $E(g, s, \Phi)$.

As we will see in Proposition 2.4, the Eisenstein series with Φ^\pm behave almost as “even/odd” functions respectively, and both have nice functional equations. This is not a coincidence. Indeed, from the point of view of representation theory, $\Phi^+(g, 0)$ is a coherent section in $I(0, \epsilon)$ in the sense that it comes from a global (two-dimensional) quadratic space, while $\Phi^-(g, 0)$ is an incoherent section in $I(0, \epsilon)$, coming from a collection of inconsistent local quadratic spaces. We refer to [Ku] for explanation of this terminology and for a general idea for computing the central derivative of incoherent Eisenstein series. Every section in $I(0, \epsilon)$ is a linear combination of coherent and incoherent sections; we just made it explicit in this case.

2. PROOF OF THEOREM 0.1

Let $\psi = \prod \psi_p$ be the “canonical” additive character of \mathbb{A} via

$$\psi_p(x) = \begin{cases} e^{2\pi i x} & \text{if } p = \infty, \\ e^{-2\pi i \lambda(x)} & \text{if } p \neq \infty. \end{cases}$$

Here λ is the canonical map $\mathbb{Q}_p \rightarrow \mathbb{Q}_p/\mathbb{Z}_p \hookrightarrow \mathbb{Q}/\mathbb{Z}$. For a standard section $\Phi = \prod \Phi_p \in I(s, \epsilon)$ and $d \in \mathbb{Q}$, one defines the local Whittaker function

$$(2.1) \quad W_{d,p}(g, s, \Phi) = \int_{\mathbb{Q}_p} \Phi(w n(b)g, s) \psi_p(-db) db.$$

Let

$$(2.2) \quad W_{d,p}^*(g, s, \Phi) = L_p(s+1, \epsilon) W_{d,p}(g, s, \Phi)$$

be its completion. We also set $M_p(s) = W_{0,p}(s)$ and $M_p^*(s) = W_{0,p}^*(s)$. So $M^*(s) = \prod M_p^*(s)$ is a normalized intertwining operator from $I(s, \epsilon)$ to $I(-s, \epsilon)$.

In general, an Eisenstein series $E^*(g, s, \Phi)$ has a Fourier expansion

$$(2.3) \quad E^*(g, s, \Phi) = \sum_d E_d^*(g, s, \Phi)$$

with

$$(2.4) \quad E_d^*(g, s, \Phi) = q^{\frac{s+1}{2}} \prod_p W_{d,p}^*(g, s, \Phi)$$

for $d \neq 0$ and

$$(2.5) \quad E_0^*(g, s, \Phi) = q^{\frac{s+1}{2}} \Lambda(s+1, \epsilon) \Phi(g, s) + q^{\frac{s+1}{2}} M^*(s) \Phi(g, s).$$

The local Whittaker integrals are computed in the next three lemmas.

Lemma 2.1 ([KRY, Lemma 2.4]). *For a finite prime number $p \neq q$, one has $W_{d,p}^*(1, s, \Phi_p) = 0$ unless $\text{ord}_p d \geq 0$. In such a case, one has*

$$W_{d,p}^*(1, s, \Phi_p) = \sum_{r=0}^{\text{ord}_p d} (\epsilon_p(p) p^{-s})^r$$

and

$$M_p^*(s) \Phi(s) = L_p(s, \epsilon) \Phi_p(-s).$$

Here Φ_p is the unique spherical section defined in section 1. In particular,

$$W_{d,p}^*(1, 0, \Phi_p) = \rho_p(d),$$

where $\rho_p(d) = \rho(p^{\text{ord}_p d})$ for $p < \infty$.

Lemma 2.2. *For $p = q$, one has*

$$\begin{pmatrix} W_{d,q}^*(w_0, s, \Phi^\pm) \\ W_{d,q}^*(w_1, s, \Phi^\pm) \end{pmatrix} = \begin{cases} (1 \pm \epsilon_q(d) q^{-s(\text{ord}_q d + 1)}) \begin{pmatrix} \pm \frac{1}{\sqrt{-q}} \\ -\frac{1}{q} \end{pmatrix} & \text{if } \text{ord}_q d \geq 0, \\ (1 \pm \epsilon_q(d)) \begin{pmatrix} 0 \\ -q^{-1} \end{pmatrix} & \text{if } \text{ord}_q d = -1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$M_q^*(s)\Phi_q^\pm = \pm \frac{1}{\sqrt{-q}} \Phi_q^\pm.$$

Proof. The first formula follows from [KRY, (3.26)-(3.29)]. For the second formula, notice that $M_q^*(s)$ is an intertwining operator between eigenspaces $W(J_q, \epsilon_q, s)$ and $W(J_q, \epsilon_q, -s)$ of J_p . So

$$M_q^*(s)\Phi_q^\pm = a^\pm \Phi_q^+ + b^\pm \Phi_q^-$$

for some constants a^\pm and b^\pm . Plugging in $g = w_0$ and w_1 , and applying the first formula, one gets the desired formula. \square

Lemma 2.3. Let $\Phi = \Phi_\infty$ be the local section in $I(s, \epsilon_\infty)$ defined by (1.3).

(1)

$$W_{d,\infty}^*(g_\tau, s, \Phi) = 2iv^{\frac{1+s}{2}} \pi^{-\frac{s}{2}} e(dv) \prod_{j=0}^k \frac{j - \frac{s}{2}}{j + \frac{s}{2}} \frac{\eta(2v, \pi d, \frac{s}{2} + k + 1, \frac{s}{2} - k)}{\Gamma(\frac{s}{2})}.$$

Here

$$\eta(g, h, \alpha, \beta) = \int_{x \pm h > 0} e^{-gx} (x+h)^{\alpha-1} (x-h)^{\beta-1} dx$$

is Shimura's eta function for $g > 0$, $h \in \mathbb{R}$, and $\operatorname{Re} \alpha$ and $\operatorname{Re} \beta$ sufficiently large [Sh].

(2) For $d > 0$, one has

$$W_{d,\infty}^*(g_\tau, 0, \Phi) = 2iv^{\frac{1}{2}} p_k(4\pi dv) e(d\tau),$$

where p_k is defined by (0.4).

(3) For $d < 0$, one has $W_{d,\infty}^*(g_\tau, 0, \Phi) = 0$, and

$$W_{d,\infty}'(g_\tau, 0, \Phi) = iv^{\frac{1}{2}} q_k(-4\pi dv) e(d\tau),$$

where q_k is given by (0.5).

(4) $M_\infty^*(s)\Phi_\infty(s) = i \prod_{j=0}^k \frac{j-s/2}{j+s/2} L_\infty(s, \epsilon) \Phi_\infty(-s)$.

Proof. The proof is the same as that of [KRY, Proposition 2.6] and is left to the reader. \square

Proposition 2.4. One has the functional equation as s goes to $-s$:

$$(2.6) \quad \prod_{j=0}^k (j - \frac{s}{2}) E^*(g, -s, \Phi^\pm) = \pm \prod_{j=0}^k (j + \frac{s}{2}) E^*(g, s, \Phi^\pm).$$

Proof. By Lemmas 2.1-2.3, one has

$$(2.7) \quad M^*(s)\Phi(g, s) = \pm q^{-\frac{1}{2}} \prod_{j=0}^k \frac{j - \frac{s}{2}}{j + \frac{s}{2}} \Lambda(s, \epsilon) \Phi(g, -s).$$

Now the proposition follows from the functional equations

$$q^{\frac{s}{2}} \Lambda(s, \epsilon) = q^{-\frac{s}{2}} \Lambda(-s, \epsilon)$$

and

$$E(g, s, \Phi) = E(g, -s, M(s)\Phi).$$

Here $M(s) = M^*(s)\Lambda(s+1, \epsilon)^{-1}$ is the unnormalized intertwining operator from $I(s, \epsilon)$ to $I(-s, \epsilon)$. \square

Theorem 2.5. *One has*

$$(2.8) \quad v^{-\frac{1}{2}} E^*(g_\tau, 0, \Phi^+) = 2(h_q + 2 \sum_{n>0} \rho(n) p_k(4\pi n v) e(n\tau))$$

and

$$(2.9) \quad \begin{aligned} v^{-\frac{1}{2}} E^{*'}(g_\tau, 0, \Phi^-) \\ = h_q(\log qv + 2 \frac{\Lambda'(1, \epsilon)}{\Lambda(1, \epsilon)} + \sum_{j=1}^k \frac{1}{j}) - 2 \sum_{n>0} a_n p_k(4\pi n v) e(n\tau) \\ - 2 \sum_{n<0} \rho(-n) q_k(-4\pi n v) e(n\tau). \end{aligned}$$

Proof. First we observe that

$$(2.10) \quad \prod_{p \nmid q\infty} \rho_p(d)(1 \pm \epsilon_q(d)) = \rho(|d|)(1 \pm \epsilon_q(d)) = 2\rho(d)$$

since

$$1 = \prod_{p \leq \infty} \epsilon_p(d) = \text{sign}(d) \epsilon_q(d) \prod_{p|d} (-1)^{\text{ord}_p d}.$$

Formula (2.8) is a special case of the Siegel-Weil formula. We give a direct proof here using Lemmas 2.1-2.3. First, the lemmas imply $E_d^*(g_\tau, 0, \Phi^+) = 0$ unless $d \geq 0$ is an integer. When $d > 0$ is an integer, the lemmas and (2.10) imply

$$(2.11) \quad \begin{aligned} E_d^*(g_\tau, 0, \Phi^+) &= q^{\frac{1}{2}} \prod_{p \nmid q\infty} \rho_p(d) \frac{1 + \epsilon_q(d)}{\sqrt{-q}} 2iv^{\frac{1}{2}} p_k(4\pi d v) e(d\tau) \\ &= 4v^{\frac{1}{2}} \rho(d) p_k(4\pi d v) e(d\tau). \end{aligned}$$

The same lemmas also imply

$$\begin{aligned} E_0^*(g_\tau, 0, \Phi^+) &= q^{\frac{1}{2}} \Lambda(1, \epsilon) \Phi^+(g_\tau, 0) + q^{\frac{1}{2}} M^*(0) \Phi^+(g_\tau, 0) \\ &= hv^{\frac{1}{2}} + \Lambda(0, \epsilon) v^{\frac{1}{2}} \\ &= 2hv^{\frac{1}{2}}. \end{aligned}$$

This proves (2.8).

As for (2.9), we again check term by term, and it is clear from the lemmas that $E_d^{*'}(g_\tau, 0, \Phi^-) = 0$ unless d is an integer, which we assume from now on.

When $d < 0$, $W_{d,\infty}^*(g_\tau, 0, \Phi^-) = 0$ by Lemma 2.3(3), and so (using Lemmas 2.1-2.3 and (2.10))

$$\begin{aligned} E_d^{*'}(g_\tau, 0, \Phi^-) &= q^{\frac{1}{2}} W_{d,\infty}^{*'}(g_\tau, 0, \Phi_\infty^-) W_{d,q}^*(1, 0, \Phi_q^-) \prod_{p \nmid q\infty} W_{d,p}^*(1, 0, \Phi_p^-) \\ &= -2v^{\frac{1}{2}} q_k(-4\pi d v) e(d\tau) (1 - \epsilon_q(d)) \prod_{p \nmid q\infty} \rho_p(d) \\ &= -2v^{\frac{1}{2}} \rho(-d) q_k(-4\pi d v) e(d\tau), \end{aligned}$$

as desired.

When $d > 0$ and $\epsilon_q(d) = 1$, one has $W_{d,q}^*(1, 0, \Phi^-) = 0$ and

$$W_{d,q}^{*'}(1, 0, \Phi^-) = \frac{-1}{\sqrt{-q}} (\text{ord}_q d + 1) \log q.$$

The same computation using Lemmas 2.1-2.3 and (2.10) yields

$$\begin{aligned} E_d^{*'}(g_\tau, 0, \Phi^-) &= -2v^{\frac{1}{2}} p_k(4\pi dv) e(d\tau) (\text{ord}_q d + 1) \rho(d) \log q \\ (2.12) \qquad \qquad \qquad &= -2v^{\frac{1}{2}} a_n p_k(4\pi dv) e(d\tau), \end{aligned}$$

since $a_n = (\text{ord}_q d + 1) \rho(d) \log q$ in this case.

When $d > 0$ and $\epsilon_q(d) = -1$, there is a prime $l|d$ such that $W_{d,l}^*(1, 0, \Phi_l) = \rho_l(d) = 0$ by (2.10). In this case,

$$W_{d,l}^{*'}(1, 0, \Phi_l) = \frac{1}{2} (\text{ord}_l d + 1) \log l.$$

The same calculation yields

$$E_d^{*'}(g_\tau, 0, \Phi^-) = -2v^{\frac{1}{2}} a_n p_k(4\pi dv) e(d\tau),$$

as desired.

Finally, when $d = 0$, one has by the same lemmas,

$$(2.13) \qquad E_0^*(g_\tau, s, \Phi^\pm) = \frac{1}{\prod_{j=1}^k (j + \frac{s}{2})} (G(s) \pm G(-s))$$

with

$$(2.14) \qquad G(s) = (qv)^{\frac{1+s}{2}} \Lambda(1+s, \epsilon) \prod_{j=1}^k (j + \frac{s}{2}).$$

So

$$(2.15) \qquad E_0^{*'}(g_\tau, 0, \Phi^-) = \frac{2G'(0)}{k!} = hv^{\frac{1}{2}} \left(\log(qv) + 2 \frac{\Lambda'(1, \epsilon)}{\Lambda(1, \epsilon)} + \sum_{j=1}^k \frac{1}{j} \right).$$

This finishes the proof of (2.9). □

Proof of Theorem 0.1. One has by Proposition 2.4,

$$(2.16) \qquad E^*(\tau, 0) = \frac{1}{2} v^{-k-\frac{1}{2}} E^*(g_\tau, 0, \Phi^+)$$

and

$$(2.17) \qquad E^{*'}(\tau, 0) = \frac{1}{2} v^{-k-\frac{1}{2}} \left[E^{*'}(g_\tau, 0, \Phi^-) - \frac{1}{2} \sum_{j=1}^k \frac{1}{j} E^*(g_\tau, 0, \Phi^+) \right].$$

Now Theorem 0.1 easily follows from Propositions 1.1 and 2.4 and Theorem 2.5.

3. PROOF OF THEOREM 0.2

By Proposition 1.1 and Formulas (2.16) and (2.17), Theorem 0.2 is equivalent to the identity

$$(3.1) \qquad \left(\frac{|\tau|}{\tau} \right)^{2k+1} \begin{pmatrix} E^*(g_{-\frac{1}{q\tau}}, 0, \Phi^+) \\ E^{*'}(g_{-\frac{1}{q\tau}}, 0, \Phi^-) \end{pmatrix} = i \begin{pmatrix} -1 & 0 \\ \frac{1}{2} \log q & 1 \end{pmatrix} \begin{pmatrix} E^*(g_\tau, 0, \Phi^+) \\ E^{*'}(g_\tau, 0, \Phi^-) \end{pmatrix}.$$

To prove (3.1), one observes the following trivial but fundamental identity and computes both sides:

$$(3.2) \qquad E^*(w_\infty^{-1} g_{q\tau}, s, \Phi^\pm) = E^*(w_f g_{q\tau}, s, \Phi^\pm).$$

Here w_f and w_∞ are the images of $w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in $G(\mathbb{A}_f)$ and $G(\mathbb{R})$ respectively. The left-hand side of this identity is given by

Lemma 3.1.

$$E^*(w_\infty^{-1}g_\tau, s, \Phi^\pm) = \left(\frac{|\tau|}{\tau}\right)^{2k+1} E^*(g_{-\frac{1}{\tau}}, s, \Phi^\pm).$$

Proof. Write $w_\infty^{-1}g_\tau = g_{-\frac{1}{\tau}}k_\theta$; then $e^{i\theta} = |\tau|/\tau$. So one has, for any $\gamma \in G(\mathbb{Q})$,

$$\Phi_\infty(\gamma_\infty w_\infty^{-1}g_\tau, s) = \left(\frac{|\tau|}{\tau}\right)^{2k+1} \Phi_\infty(\gamma_\infty g_{-\frac{1}{\tau}}, s).$$

Plugging this into the definition of the Eisenstein series, one gets the lemma. \square

For the right-hand side of (3.2), one has

Lemma 3.2.

$$\begin{pmatrix} E^*(w_f g_{q\tau}, 0, \Phi^+) \\ E^{*'}(w_f g_{q\tau}, 0, \Phi^-) \end{pmatrix} = i \begin{pmatrix} -1 & 0 \\ \frac{1}{2} \log q & 1 \end{pmatrix} \begin{pmatrix} E^*(g_\tau, 0, \Phi^+) \\ E^{*'}(g_\tau, 0, \Phi^-) \end{pmatrix}.$$

Proof. We verify these identities by comparing the Fourier coefficients $E_{\frac{d}{q}}^*(w_f g_{q\tau}, s, \Phi^\pm)$ with $E_d^*(g_\tau, s, \Phi^\pm)$. We may assume that d is an integer by Lemmas 2.1-2.3. Straightforward calculation using the same lemmas yields, for any integer d ,

$$(3.3) \quad W_{\frac{d}{q}, p}^*(w_f g_{q\tau}, s, \Phi^\pm) = F_p^\pm(d) W_{d, p}^*(g_\tau, s, \Phi^\pm)$$

with

$$(3.4) \quad F_p^\pm(d) = \begin{cases} 1 & \text{if } p \nmid q\infty, \\ q^{\frac{1-s}{2}} & \text{if } p = \infty, \\ \pm \frac{1}{\sqrt{-q}} \frac{1 \pm \epsilon_q(d) q^{-sr}}{1 \pm \epsilon_q(d) q^{-s(r+1)}} & \text{if } p = q. \end{cases}$$

Here $r = \text{ord}_q d$. We will verify the derivative part and leave the value part to the reader. First assume $d \neq 0$. It follows from (3.3) that

$$E_{\frac{d}{q}}^{*'}(w_f g_{q\tau}, 0, \Phi^-) = i E_d^{*'}(g_\tau, 0, \Phi^-) \begin{cases} 1 & \text{if } \epsilon_q(d) = -1, \\ 1 - \frac{1}{\text{ord}_q d + 1} & \text{if } \epsilon_q(d) = 1. \end{cases}$$

When $\epsilon_q(d) = 1$, one has by (2.11) and (2.12),

$$E_d^{*'}(g_\tau, 0, \Phi^-) = -E_d^*(g_\tau, 0, \Phi^+) \frac{\text{ord}_q d + 1}{2} \log q.$$

So

$$(3.5) \quad E_{\frac{d}{q}}^{*'}(w_f g_{q\tau}, 0, \Phi^-) = i E_d^{*'}(g_\tau, 0, \Phi^-) + \frac{i}{2} \log q E_d^*(g_\tau, 0, \Phi^+),$$

as desired. When $\epsilon_q(d) = -1$ we have $E_d^*(g_\tau, 0, \Phi^+) = 0$, and (3.5) still holds.

It remains to check the constant term. Recall (2.13)-(2.15). Direct calculation using Lemmas 2.1-2.3 also gives

$$(3.6) \quad E_0^*(w_f g_{q\tau}, s, \Phi^\pm) = \mp \frac{i}{\prod_{j=1}^k (j + \frac{s}{2})} (q^{\frac{s}{2}} G(s) \pm q^{-\frac{s}{2}} G(-s)).$$

Therefore,

$$\begin{aligned} E_0^{*'}(w_f g_{q\tau}, 0, \Phi^-) &= i \frac{2G'(0)}{k!} + i \frac{2G(0)}{k!} \frac{1}{2} \log q \\ (3.7) \qquad \qquad \qquad &= i E_0^{*'}(g_\tau, 0, \Phi^-) + \frac{i}{2} \log q E_0^*(g_{q\tau}, 0, \Phi^+), \end{aligned}$$

as expected, too. □

4. *L*-SERIES

Recall that q is a prime congruent to 3 modulo 4 and $\mathbf{k} = \mathbb{Q}(\sqrt{-q})$ is the associated imaginary quadratic field. Recall also ([Roh]) that a canonical Hecke character of \mathbf{k} of weight $2k + 1$ is a Hecke character μ satisfying

- (1) The conductor of μ is $\sqrt{-q}\mathcal{O}_{\mathbf{k}}$.
- (2) $\mu(\bar{\mathfrak{A}}) = \overline{\mu(\mathfrak{A})}$ for an ideal \mathfrak{A} relatively prime to $\sqrt{-q}\mathcal{O}_{\mathbf{k}}$.
- (3) $\mu(\alpha\mathcal{O}_{\mathbf{k}}) = \pm \alpha^{2k+1}$.

In this section, we will give an explicit formula for the central derivative of its L -function, which has deep arithmetic implications as mentioned in the introduction. We refer to [Gro] for the arithmetics of elliptic curves associated to these Hecke characters (see also [MY] and [Ya] and the reference there for more recent developments). For each ideal class C of \mathbf{k} , we can define the partial L -series by

$$(4.1) \qquad L(s, \mu, C) = \sum_{\mathfrak{B} \in C, \text{ integral}} \mu(\mathfrak{B})(N\mathfrak{B})^{-s}.$$

Of course, $L(s, \mu) = \sum_{C \in \text{CL}(\mathbf{k})} L(s, \mu, C)$. The following proposition is standard.

Proposition 4.1. *Let $\mathfrak{A} \in C$ be a primitive ideal of \mathbf{k} relatively prime to $2q$, and write*

$$(4.2) \qquad \mathfrak{A} = [a, \frac{b + \sqrt{-q}}{2}], \quad \text{with } a > 0, \ b \equiv 0 \pmod{q}.$$

Let $\tau_{\mathfrak{A}} = \frac{b + \sqrt{-q}}{2aq}$. Then

$$L(s + k + 1, \mu, C) = \frac{\mu(\mathfrak{A})}{(N\mathfrak{A})^{2k+1}} (2\sqrt{q})^{s-k} L(2s + 1, \epsilon) E(\tau_{\mathfrak{A}}, 2s).$$

Set

$$(4.3) \qquad \theta_k(\tau) = h + 2 \sum_{n>0} \rho(n) p_k(4\pi n v) e(n\tau)$$

and

$$(4.4) \qquad \phi_k(\tau) = a_0(v) - 2 \sum_{n>0} a_n p_k(4\pi n v) e(n\tau) - 2 \sum_{n<0} \rho(-n) q_k(-4\pi n v) e(n\tau).$$

Then Theorem 0.1 says that

$$(4.5) \qquad E^*(\tau, 0) = v^{-k} \theta_k(\tau)$$

and

$$(4.6) \qquad E^{*'}(\tau, 0) = \frac{1}{2} v^{-k} (\phi_k(\tau) - \frac{1}{2} \sum_{j=1}^k \frac{1}{j} \theta_k(\tau)).$$

Corollary 4.2. *Let the notation be as in Proposition 4.1.*

(1) *The central L -value is*

$$L(k+1, \mu, C) = \frac{\pi\mu(\mathfrak{A})}{\sqrt{q}(N\mathfrak{A})^{k+1}} \theta_k(\tau_{\mathfrak{A}}).$$

(2) *When the root number of μ is -1 , i.e., $(-1)^k(\frac{2}{q}) = -1$, the central L -derivative*

$$L'(k+1, \mu, C) = \frac{\pi\mu(\mathfrak{A})}{\sqrt{q}(N\mathfrak{A})^{k+1}} \phi_k(\tau_{\mathfrak{A}}).$$

In particular,

$$\begin{aligned} & L'(k+1, \mu, \text{trivial}) \\ &= \frac{\pi}{\sqrt{q}} \phi_k\left(\frac{1}{2} + \frac{i}{2\sqrt{q}}\right) \\ &= \frac{\pi}{\sqrt{q}} \left[h\left(\log \frac{\sqrt{q}}{2}\right) + 2 \frac{\Lambda'(1, \epsilon)}{\Lambda(1, \epsilon)} + \sum_{j=1}^k \frac{1}{j} \right. \\ &\quad \left. - 2 \sum_{n>0} (-1)^n a_n p_k\left(\frac{2\pi n}{\sqrt{q}}\right) e^{-\frac{\pi n}{\sqrt{q}}} - 2 \sum_{n<0} (-1)^n \rho(-n) q_k\left(-\frac{2\pi n}{\sqrt{q}}\right) e^{-\frac{\pi n}{\sqrt{q}}} \right]. \end{aligned}$$

Proof. Only the second one needs a little explanation. When $(-1)^k(\frac{2}{q}) = -1$ we have $L(k+1, \mu, C) = 0$ automatically and thus $\theta_k(\tau_{\mathfrak{A}}) = 0$. So Theorem 0.1 and Proposition 4.1 imply

$$\begin{aligned} L'(k+1, \mu, C) &= \frac{\pi\mu(\mathfrak{A})}{\sqrt{q}(N\mathfrak{A})^{2k+1}} (2\sqrt{q})^{-k} 2E^{*'}(\tau_{\mathfrak{A}}, 0) \\ &= \frac{\pi\mu(\mathfrak{A})}{\sqrt{q}(N\mathfrak{A})^{k+1}} (\phi_k(\tau_{\mathfrak{A}}) - \frac{1}{2} \sum_{j=1}^k \frac{1}{j} \theta_k(\tau_{\mathfrak{A}})) \\ &= \frac{\pi\mu(\mathfrak{A})}{\sqrt{q}(N\mathfrak{A})^{k+1}} \phi_k(\tau_{\mathfrak{A}}). \end{aligned}$$

When C is trivial, one can take $\mathfrak{A} = \mathcal{O}_k$. In this case, $a = 1$ and $\frac{b}{2q} \equiv \frac{1}{2} \pmod{1}$, and thus

$$\phi_k(\tau_{\mathfrak{A}}) = \phi_k\left(\frac{1}{2} + \frac{i}{2\sqrt{q}}\right).$$

□

In recent joint work with S. Miller ([MY]), we proved that $L'(1, \mu, \text{trivial}) > 0$ when $q \equiv 3 \pmod{8}$ and $k = 0$. Combining that with Corollary 4.2, one has the following curious inequality:

$$\begin{aligned} (4.7) \quad & h\left(\log \frac{\sqrt{q}}{2}\right) + 2 \frac{\Lambda'(1, \epsilon)}{\Lambda(1, \epsilon)} + \sum_{j=1}^k \frac{1}{j} \\ & > 2 \sum_{n>0} (-1)^n a_n p_k\left(\frac{2\pi n}{\sqrt{q}}\right) e^{-\frac{\pi n}{\sqrt{q}}} + 2 \sum_{n<0} (-1)^n \rho(-n) q_k\left(-\frac{2\pi n}{\sqrt{q}}\right) e^{-\frac{\pi n}{\sqrt{q}}}. \end{aligned}$$

ACKNOWLEDGEMENT

This work was inspired by joint work with Steve Kudla and Michael Rapoport. The author thanks them for the inspiration. He thanks Rene Schoof for numerically verifying the formulae in Corollary 4.2, which corrects a mistake in an earlier version of this paper. Finally, he thanks Dick Gross, Steve Miller, and David Rohrlich for stimulating discussions.

REFERENCES

- [Gro] B. Gross, *Arithmetic on elliptic curves with complex multiplication*, Lecture Notes in Math., no. 776, Springer-Verlag, Berlin, 1980. MR **81f**:10041
- [GZ] B. Gross and D. Zagier, *Heegner points and derivatives of L -series*, Invent. Math. **84** (1986), 225-320. MR **87j**:11057
- [Ku] S. Kudla, *Central derivatives of Eisenstein series and height pairings*, Ann. Math. **146** (1997), 545-646. MR **99j**:11047
- [KRY] S. Kudla, M. Rapoport, and T.H. Yang, *On the derivative of an Eisenstein series of weight one*, Internat. Math. Res. Notices **7** (1999), 347-385. MR **2000b**:11057
- [MY] S. D. Miller and T. H. Yang, *Non-vanishing of the central derivative of canonical Hecke L -functions*, Math. Res. Letters **7** (2000), 263-277. MR **2001i**:11058
- [Roh] D. Rohrlich, *Root numbers of Hecke L -functions of CM fields*, Amer. J. Math. **104** (1982), 517-543. MR **83j**:12011
- [Sh] G. Shimura, *Confluent hypergeometric functions on tube domains*, Math. Ann. **260** (1982), 269-302. MR **84f**:32040
- [Ya] T.H. Yang, *On CM abelian varieties over imaginary quadratic fields*, preprint.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN 53706
E-mail address: thyang@math.wisc.edu

SYSTEMS OF DIAGONAL DIOPHANTINE INEQUALITIES

ERIC FREEMAN

ABSTRACT. We treat systems of real diagonal forms $F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_R(\mathbf{x})$ of degree k , in s variables. We give a lower bound $s_0(R, k)$, which depends only on R and k , such that if $s \geq s_0(R, k)$ holds, then, under certain conditions on the forms, and for any positive real number ϵ , there is a nonzero integral simultaneous solution $\mathbf{x} \in \mathbb{Z}^s$ of the system of Diophantine inequalities $|F_i(\mathbf{x})| < \epsilon$ for $1 \leq i \leq R$. In particular, our result is one of the first to treat systems of inequalities of even degree. The result is an extension of earlier work by the author on quadratic forms. Also, a restriction in that work is removed, which enables us to now treat combined systems of Diophantine equations and inequalities.

1. INTRODUCTION

1.1. Statement of main result. In 1980, Schmidt [17] proved a far-reaching result about systems of Diophantine inequalities of odd degree. Given any odd positive integers d_1, \dots, d_R , Schmidt showed that there exists a positive integer $s_1 = s_1(d_1, \dots, d_R)$, depending only on d_1, \dots, d_R , with the following property: given any positive integer $s \geq s_1$ and any real forms, or homogeneous polynomials, $G_1(\mathbf{x}), \dots, G_R(\mathbf{x})$, in s variables, of respective degrees d_1, \dots, d_R , and given any positive number ϵ , there always exists a nonzero integral vector $\mathbf{y} \in \mathbb{Z}^s$ satisfying the system

$$(1.1) \quad |G_1(\mathbf{y})| < \epsilon, |G_2(\mathbf{y})| < \epsilon, \dots, |G_R(\mathbf{y})| < \epsilon.$$

So, in other words, as long as the forms are all of odd degree, and are defined in enough variables in terms only of the degrees, then there is a nonzero integral solution of the inequalities (1.1). Many particular classes of systems of the type (1.1) have been studied.

For Diophantine inequalities of even degree, the situation is much different. There is no such general result as above for integral solutions of Diophantine inequalities of even degree, and in fact there are few results at all for inequalities of even degree. (However, results are known if one allows solutions in algebraic integers in purely imaginary number fields. See Theorem 11.1 of [22].) In this article, we present one of the first results concerning systems of Diophantine inequalities of even degree, while at the same time removing a restriction from an earlier paper by the author, on quadratic Diophantine inequalities [10]. We are now able to remove

Received by the editors October 15, 2001.

2000 *Mathematics Subject Classification.* Primary 11D75; Secondary 11D41, 11D72, 11P55.

Key words and phrases. Combined systems of Diophantine equations and inequalities, forms in many variables, applications of the Hardy-Littlewood method.

The author was supported by an NSF Postdoctoral Fellowship.

the restriction by combining the powerful ideas of Bentkus and Götze [3] with the techniques of Nadesalingam and Pitman [16], and by adapting our previous work in [10] and [11] to treat the minor arcs properly.

To state our first result, we require some notation and definitions. We shall be working with systems of diagonal forms $F_i(\mathbf{x})$ given by

$$(1.2) \qquad F_i(\mathbf{x}) = \lambda_{i1}x_1^k + \lambda_{i2}x_2^k + \dots + \lambda_{is}x_s^k \quad (1 \leq i \leq R).$$

For systems of forms F_i as in (1.2), we define the coefficient matrix of the system \mathbf{F} to be the matrix

$$(1.3) \qquad A = (\lambda_{ij})_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}}.$$

For $1 \leq j \leq s$, we denote the j^{th} column of A by $\boldsymbol{\lambda}_j$.

Now suppose that J is a subset of the set of indices $\{1, 2, \dots, s\}$. We define A_J to be the submatrix of A consisting of the columns $\boldsymbol{\lambda}_j$ with $j \in J$, and we define $r(A_J)$ to be the rank of the matrix A_J . Finally, if $\mathbf{x} \in \mathbb{R}^s$ satisfies $F_i(\mathbf{x}) = 0$ for $1 \leq i \leq R$ and the matrix

$$\left(\frac{\partial F_i}{\partial x_j}\right)_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}}$$

is of full rank, then we say that \mathbf{x} is a nonsingular solution of the system \mathbf{F} .

Now, for integers R and k and any real number u , we define the functions

$$(1.4) \qquad m_0(R, k, u) = \begin{cases} \min(4R^2 + 4R + 1, 384 \log 16R + 5) & \text{if } k = 2 \\ \frac{Rk \log 2k}{\log 2} + uk \log(R \log 2k) & \text{if } k \text{ is odd and } k \geq 3 \\ k [48k^2 \log 3Rk^2] & \text{if } k \geq 3, \end{cases}$$

and

$$(1.5) \qquad n_0(k, u) = \begin{cases} 5 & \text{if } k = 2 \\ \min\left(2^k + 1, k(\log k + \log \log k + 2) + \frac{uk \log \log k}{\log k}\right) & \text{if } k \geq 3. \end{cases}$$

We can now state our first result.

Theorem 1.1. *Suppose that k is an integer with $k \geq 2$ and that R is a positive integer. There are absolute real positive constants \tilde{C}_1 and \tilde{C}_2 for which the following property holds:*

Suppose that ℓ is an integer satisfying

$$(1.6) \qquad \ell \geq \max\left(m_0\left(R, k, \tilde{C}_1\right), n_0\left(k, \tilde{C}_2\right)\right).$$

Suppose that s is an integer satisfying $s \geq \ell R$. For $1 \leq i \leq R$, suppose that $F_i(\mathbf{x})$ is a real diagonal form of degree k , as in (1.2). Let A be the coefficient matrix of the system \mathbf{F} , as in (1.3). Assume that the following two conditions are satisfied:

(i) *Either k is odd, or there exists a real nonsingular solution \mathbf{y} of the system*

$$F_1(\mathbf{y}) = F_2(\mathbf{y}) = \dots = F_R(\mathbf{y}) = 0.$$

(ii) *For every subset $J \subseteq \{1, 2, \dots, s\}$, one has $|J| \leq s - \ell(R - r(A_J))$.*

Fix any positive real number ϵ . Then there is a nonzero integral solution $\mathbf{x} \in \mathbb{Z}^s$ of the system

$$(1.7) \qquad |F_i(\mathbf{x})| < \epsilon \quad \text{for } 1 \leq i \leq R.$$

For general even k , Theorem 1.1 is one of the first results of its kind to our knowledge. We note that at least some conditions similar to (i) and (ii) are necessary, as may be seen by considering the examples given after Theorem 2 of [7].

For odd k , we note that Theorem 1.1 is not very much of an improvement beyond that given by Nadesalingam and Pitman [16], and could presumably be obtained by combining their methods with results of Vaughan [19], [20] and work of Wooley [23]. We also observe that, following the method of Section 7.2 of [16], we could remove condition (ii) for odd k if we chose to do so.

As well, our method of proof shows that under the conditions of Theorem 1.1, we can give a lower bound of the expected order of magnitude, P^{s-Rk} , for the number of solutions of (1.7) in a box of size P , for all sufficiently large P . This was not previously known, even in the special case of systems of inequalities of odd degree. We emphasize that when using our methods, condition (ii) is necessary to obtain this lower bound. Presumably, one could also give an asymptotic formula for the number of solutions, by combining with the methods of [12].

Note that we have excluded the case $k = 1$ from the statement of the theorem. In this case, our knowledge is much better. For $k = 1$, if one has $s \geq R + 1$, a nonzero solution of (1.7) may be found using a box principle, whether or not condition (ii) holds. (See the Lemma in [4].) For $k = 1$, one can also find many solutions of (1.7) in a box of size P . (See Lemma 1 of [9].)

1.2. Combined systems of Diophantine equations and inequalities. We note now that it would actually be fairly routine to give at least one result on inequalities of even degree; one could simply generalize the work in [10] by combining those techniques with the methods in [11]. However, such a generalization would exclude many important classes of systems of inequalities, for example those in which some of the forms are integral. We were forced to exclude such systems in [10] because of the methods we used. In our current work, we are able to treat these formerly excluded systems. We give some background to more fully explain.

In [10], we considered simultaneous systems of diagonal quadratic Diophantine inequalities. For a positive integer R , define, for $1 \leq i \leq R$, the real quadratic forms

$$Q_i(\mathbf{x}) = \lambda_{i1}x_1^2 + \lambda_{i2}x_2^2 + \dots + \lambda_{is}x_s^2.$$

It was proved in [10] that for every positive real number ϵ , under certain conditions on the system of forms Q_1, Q_2, \dots, Q_R , there is an integral vector $\mathbf{x} \in \mathbb{Z}^s \setminus \{\mathbf{0}\}$ such that one has

$$|Q_i(\mathbf{x})| < \epsilon \quad (1 \leq i \leq R).$$

In that paper, one of the conditions we assumed was the following. (See condition (iii) of Theorem 3 of [10].)

For each choice of $(\beta_1, \beta_2, \dots, \beta_R) \in \mathbb{R}^R \setminus \{\mathbf{0}\}$, there is at least

(1.8) one coefficient of $\beta_1 Q_1 + \beta_2 Q_2 + \dots + \beta_R Q_R$ that is irrational.

This condition allowed us to use a modification of the remarkable work of Bentkus and Götze [3], but excludes certain important systems from consideration. The restriction (1.8) rules out systems in which one or more of the forms is an integral form, and also any system in which any nontrivial linear combination of the forms

is an integral form. The central new contribution of this paper is in removing the condition (1.8).

We now state a more technical and more general version of Theorem 1.1. We require more notation. For a real vector $\beta = (\beta_1, \dots, \beta_R) \in \mathbb{R}^R$ and a system \mathbf{G} of forms $G_1(\mathbf{x}), G_2(\mathbf{x}), \dots, G_R(\mathbf{x})$, we define the form

$$(\beta \cdot \mathbf{G})(\mathbf{x}) = \beta_1 G_1(\mathbf{x}) + \beta_2 G_2(\mathbf{x}) + \dots + \beta_R G_R(\mathbf{x}).$$

Also, for real numbers x , we define

$$e(x) = e^{2\pi i x}.$$

Theorem 1.2. *Suppose that k is an integer with $k \geq 2$ and that r and R are integers with $R \geq 1$ and $0 \leq r \leq R$. Then there are absolute positive real constants \tilde{C}_1 and \tilde{C}_2 with the following property:*

Define $n_0(k, u)$ as in (1.5). Suppose that ℓ is an integer satisfying

$$(1.9) \quad \ell \geq n_0(k, \tilde{C}_2).$$

Let s be an integer with $s \geq \ell R$. Also suppose, for $1 \leq i \leq R$, that

$$F_i(\mathbf{x}) = \lambda_{i1}x_1^k + \lambda_{i2}x_2^k + \dots + \lambda_{is}x_s^k$$

is a diagonal form with real coefficients. Let A be the coefficient matrix of the system \mathbf{F} , as in (1.3).

Assume that the following four conditions are satisfied:

(i) *Either k is odd, or there exists a real nonsingular solution \mathbf{y} of the system*

$$F_1(\mathbf{y}) = F_2(\mathbf{y}) = \dots = F_R(\mathbf{y}) = 0.$$

(ii) *For every subset $J \subseteq \{1, 2, \dots, s\}$, one has $|J| \leq s - \ell(R - r(A_J))$.*

(iii) *The forms F_1, F_2, \dots, F_r have integer coefficients; also, if $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_R) \in \mathbb{R}^R$ and $\alpha \cdot \mathbf{F}$ is a rational form, then $\alpha_{r+1} = \alpha_{r+2} = \dots = \alpha_R = 0$.*

(iv) *If $r \geq 1$ holds, then there is a positive real constant $c(\mathbf{F})$ such that one has*

$$\mathfrak{S} = \sum_{q=1}^{\infty} \sum_{\substack{\mathbf{a}: (a_1, \dots, a_r, q)=1 \\ 1 \leq a_i \leq q \ (1 \leq i \leq r)}} q^{-s} \prod_{j=1}^s \sum_{x=1}^q e\left(\frac{x^k}{q} \sum_{i=1}^r \lambda_{ij} a_i\right) \geq c(\mathbf{F}).$$

Fix any positive real number ϵ . Then there is a nonzero integral solution $\mathbf{x} \in \mathbb{Z}^s$ of the system

$$(1.10) \quad \begin{aligned} F_i(\mathbf{y}) &= 0 & \text{for } 1 \leq i \leq r, \\ |F_i(\mathbf{y})| &< \epsilon & \text{for } r+1 \leq i \leq R. \end{aligned}$$

Moreover, if we define $m_0(r, k, u)$ as in (1.4) and we assume that the condition $\ell \geq m_0(r, k, \tilde{C}_1)$ holds, then we may omit condition (iv) from our assumptions.

Some discussion of the condition (iii) is warranted here, since it is the most important distinction between Theorems 1.1 and 1.2. In Theorem 1.1, we consider systems of inequalities (1.7). Now, if one chooses $\epsilon < 1$, and $F_1(\mathbf{x})$, say, is actually an integral form, then the system (1.7) reduces to the system

$$\begin{aligned} F_1(\mathbf{y}) &= 0, \\ |F_i(\mathbf{y})| &< \epsilon & \text{for } 2 \leq i \leq R. \end{aligned}$$

So, in this case, the system of R inequalities actually reduces to a system of one equation and $R - 1$ inequalities. A similar reduction occurs if some nontrivial real linear combination of the forms F_i , say $\alpha_1 F_1 + \alpha_2 F_2 + \dots + \alpha_R F_R$, is an integral form. In these situations, one might say that there is actually an equation hidden in the system of inequalities. The condition (iii) ensures that there are actually r equations in the system and $R - r$ "true" inequalities. It is helpful to ensure that there are not any more "hidden" equations because it turns out that one requires more variables to treat the system if there are more equations present. One may think of Theorem 1.2, more or less, as the sub-case of Theorem 1.1 in which there are exactly r equations present in a system of R inequalities. We note that the second clause of condition (iii) is vacuous if $r = R$ holds.

One might question if, in condition (iii), the term $\alpha \cdot \mathbf{F}$ could be replaced by the term $\alpha_{r+1} F_{r+1} + \alpha_{r+2} F_{r+2} + \dots + \alpha_R F_R$, to give a slightly weaker condition asserting that r equations are "hidden" in the system. It turns out that one can not, as may be seen by considering the example

$$\begin{aligned} F_1(\mathbf{x}) &= 2x_1^k + 3x_2^k + a_3x_3^k + a_4x_4^k + \dots + a_sx_s^k, \\ F_2(\mathbf{x}) &= (2 + \sqrt{2})x_1^k + (3 + \sqrt{2})x_2^k + (a_3 + \sqrt{2})x_3^k \\ &\quad + (a_4 + \sqrt{2})x_4^k + \dots + (a_s + \sqrt{2})x_s^k, \end{aligned}$$

where a_3, a_4, \dots, a_s are any integers. Here, for $r = 1$ and $R = 2$, condition (iii) does not hold, since $(1/\sqrt{2})(F_2 - F_1)$ is an integral form, and thus the system (1.10) is equivalent in this case, for small ϵ , to the system

$$F_1(\mathbf{x}) = \frac{1}{\sqrt{2}}(F_2(\mathbf{x}) - F_1(\mathbf{x})) = 0;$$

on the other hand, for any nonzero real number α_2 , the form $\alpha_2 F_2$ is not a rational form since the ratio $\left((3 + \sqrt{2}) / (2 + \sqrt{2})\right)$ is irrational; so this system does not satisfy the suggested replacement condition. So the putative replacement condition is not strong enough.

We now discuss condition (iv). The term \mathfrak{S} is the so-called singular series, and condition (iv) simply states that it is bounded below by a positive constant, a necessary precondition when using the Hardy-Littlewood method. As the last sentence of Theorem 1.2 states, we could have omitted the condition from our assumptions in favor of a lower bound for ℓ . However, since the consideration of the singular series is not our central focus in this work, we have chosen to include condition (iv) so that our result can be improved immediately and transparently if improvements arise concerning the singular series and the p -adic problem. This should certainly be possible in the case $k = 2$, for example. Also, we wish to clearly indicate that the condition $\ell \geq m_0(r, k, \tilde{C}_1)$ is needed only because of the p -adic problem.

1.3. Related results. We now compare our work with other results. For even k , Theorem 1.1 is an analogue of a result of Davenport and Lewis, concerning systems of Diophantine equations of even degree. (See Theorem 2 of [7].) They assume that

a system of diagonal equations

$$F_1(\mathbf{x}) = F_2(\mathbf{x}) = \cdots = F_R(\mathbf{x}) = 0,$$

of even degree k , with $k > 2$, has a real nonsingular solution and also that for $1 \leq S \leq R$, every set of S independent integral linear combinations of F_1, \dots, F_R contains at least

$$(1.11) \quad [48RSk^2 \log(3Rk^2)]$$

variables that appear explicitly. Under these conditions, the system of equations has a nonzero integral solution. We note that if one replaces the quantity in (1.11) by $S\ell$, and restricts to integral forms, then one can show that their second condition is equivalent to condition (ii) of Theorem 1.1.

Nadesalingam and Pitman [16] proved that any R real diagonal Diophantine inequalities of odd degree k , with $k \geq 13$, in s variables with

$$s \geq 3R^2k^2 \log(3Rk)$$

have a nonzero solution. We note that they do not require any condition that is similar to condition (ii) of Theorem 1.1. Also, we observe that they could certainly have used their methods to obtain similar results, although with a different lower bound for s , in the cases $k < 13$, but in order to streamline the presentation they did not do so.

Finally, we note that Brüdern and Cook [5] have given a result on systems of diagonal Diophantine inequalities of odd degree. Under certain conditions on the coefficient matrix of the system, they show that there is a nonzero solution of the system of inequalities. They require an assumption similar to condition (ii) of Theorem 1.1 and also a condition that is stronger than (1.8). The number of variables they require is on the order of $Rn_0 \left(k, \tilde{C}_2\right)$. We also note that they can find a lower bound of the expected order of magnitude for the number of solutions of their system in a box of size P for a sequence of positive P tending to infinity, although not for all large P , as our treatment provides.

1.4. Methods used. The general strategy of the proof is to combine the method of Bentkus and Götze [3], which is very effective for Diophantine inequalities, with the techniques that Nadesalingam and Pitman [16] use to treat combined systems of Diophantine equations and inequalities. We remark that the techniques of Nadesalingam and Pitman are themselves a combination of the Hardy-Littlewood method and the Davenport-Heilbronn method. Using the techniques of Nadesalingam and Pitman allows us to treat those systems of inequalities that contain “hidden” equations. For those who are familiar with their argument, we note that we do not have a so-called residual set in our proof, as in their paper.

One other crucial result is needed, and this involves showing that, on the minor arcs, our exponential sums are smaller than the trivial bound. In previous work on these types of problems, including [10], [11], and essentially also [3], such a result was achieved by splitting the minor arcs into two regions and handling each separately. In this paper, we handle both of these regions together, which is not only cleaner, but also seems to be necessary here.

I would like to thank Scott Parsell for showing me how to improve Lemma 6.1. I would also like to thank Michael Knapp and Professor Wooley for indicating to me how to prove part of Lemma 8.5.

2. DEDUCTION OF THEOREM 1.1 FROM THEOREM 1.2

The bulk of this paper is dedicated to proving Theorem 1.2. In this section, however, we demonstrate how Theorem 1.2 implies Theorem 1.1. To this end, we consider a system \mathbf{F} of real diagonal forms F_1, F_2, \dots, F_R as in Theorem 1.1.

We give a definition first. Suppose that G_1, G_2, \dots, G_R and H_1, H_2, \dots, H_R are two systems of forms. If there exists a set of R linearly independent real vectors $\beta_1, \beta_2, \dots, \beta_R \in \mathbb{R}^R$ such that

$$H_i(\mathbf{x}) = \beta_i \cdot \mathbf{G} \quad \text{for } 1 \leq i \leq R,$$

then we say that the system \mathbf{H} is equivalent to the system \mathbf{G} , which we denote by $\mathbf{G} \sim \mathbf{H}$. It is easy to check that this is in fact an equivalence relation. We observe as well that if \mathbf{G} is, in particular, a system of diagonal forms, and $\mathbf{G} \sim \mathbf{H}$ holds, then \mathbf{H} is also a system of diagonal forms.

For any system of forms \mathbf{G} , we define $z(\mathbf{G})$ to be the number of forms among G_1, G_2, \dots, G_R that are integral, that is, whose coefficients are all integers. Now for our system of forms \mathbf{F} , we define

$$r = r(\mathbf{F}) = \max_{\mathbf{G} \sim \mathbf{F}} z(\mathbf{G}).$$

In other words, $r(\mathbf{F})$ is simply the maximum number of forms that are integral in any system \mathbf{G} equivalent to \mathbf{F} . We clearly have $0 \leq r \leq R$.

Now suppose that \mathbf{G} is a system equivalent to \mathbf{F} and that \mathbf{G} has r integral forms. So there exist R real linearly independent vectors $\beta_1, \beta_2, \dots, \beta_R \in \mathbb{R}^R$ such that $G_i = \beta_i \cdot \mathbf{F}$ for $1 \leq i \leq R$, and also the system \mathbf{G} contains r integral forms. By relabeling if necessary, we may assume that G_1, G_2, \dots, G_r are integral forms. We now show that conditions (i)-(iv) of Theorem 1.2 hold for this system \mathbf{G} . Then we will apply Theorem 1.2 to \mathbf{G} , and we will see that the nonzero solution of the system \mathbf{G} is also, under certain conditions, a solution of the system \mathbf{F} .

Since \mathbf{F} is equivalent to \mathbf{G} , if the coefficient matrix of \mathbf{F} is A , then the coefficient matrix of \mathbf{G} is TA for the nonsingular $R \times R$ matrix T with rows β_i . Thus, for any subset $J \subseteq \{1, 2, \dots, s\}$, we have

$$(2.1) \quad r((TA)_J) = r(TA_J) = r(A_J).$$

The system \mathbf{F} has a nonsingular solution \mathbf{x} if and only if there is a subset J of $\{1, 2, \dots, s\}$ with $|J| = R$ satisfying $r(A_J) = R$ and $\prod_{j \in J} x_j \neq 0$. Thus the existence

of a real nonsingular solution for \mathbf{G} follows from the existence of such a solution for \mathbf{F} . So condition (i) holds for \mathbf{G} . By (2.1), it is easy to see that condition (ii) holds for the coefficient matrix of \mathbf{G} , because it holds for the coefficient matrix of \mathbf{F} .

Now we turn to showing that condition (iii) of Theorem 1.2 holds for the system \mathbf{G} . To this end, suppose that $\alpha' = (\alpha'_1, \alpha'_2, \dots, \alpha'_R) \in \mathbb{R}^R$ is a real vector such that $\alpha' \cdot \mathbf{G}$ is a rational form. We need to show that $\alpha'_{r+1} = \alpha'_{r+2} = \dots = \alpha'_R = 0$ holds. This holds vacuously if $r = R$. For $r < R$, clearing denominators, we see that there is a nonzero integer n such that, defining $\alpha = n\alpha'$, we have

$$\alpha \cdot \mathbf{G} = n\alpha' \cdot \mathbf{G} \in \mathbb{Z}[\mathbf{x}].$$

Since n is nonzero, to prove that $\alpha'_{r+1} = \alpha'_{r+2} = \dots = \alpha'_R = 0$ holds, it is enough to prove that we have $\alpha_{r+1} = \alpha_{r+2} = \dots = \alpha_R = 0$.

So suppose that this is not the case. Letting $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_R$ be the standard unit basis for \mathbb{R}^R , we then have that $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r, \alpha$ are $r+1$ linearly independent

vectors in \mathbb{R}^R . We may thus extend this set to a basis, say $\gamma_1, \gamma_2, \dots, \gamma_R$ of \mathbb{R}^R , with $\gamma_i = \mathbf{e}_i$ for $1 \leq i \leq r$, and $\gamma_{r+1} = \boldsymbol{\alpha}$. Then the system of forms $\gamma_1 \cdot \mathbf{G}, \gamma_2 \cdot \mathbf{G}, \dots, \gamma_R \cdot \mathbf{G}$ is equivalent to \mathbf{G} , and thus in turn to \mathbf{F} . But its first $r+1$ forms are integral. This contradicts the definition of $r(\mathbf{F})$, and thus we must in fact have

$$\alpha_{r+1} = \alpha_{r+2} = \dots = \alpha_R = 0.$$

Thus condition (iii) holds for \mathbf{G} .

Now, since ℓ satisfies (1.6), we have $\ell \geq m_0(R, k, \tilde{C}_1)$, and thus we certainly have $\ell \geq m_0(r, k, \tilde{C}_1)$, whence by the final sentence of Theorem 1.2, condition (iv) is unnecessary. Since we have also seen that conditions (i)-(iii) of Theorem 1.2 hold, we may apply Theorem 1.2 to the system \mathbf{G} .

Before doing so, we give some more notation. For a vector $\mathbf{x} = (x_1, x_2, \dots, x_s) \in \mathbb{R}^s$, define

$$|\mathbf{x}| = \max_{1 \leq j \leq s} |x_j|.$$

For an $R \times s$ matrix $M = (m_{ij})_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}}$, we define

$$\|M\| = \max_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}} |m_{ij}|.$$

We note that the notation differs slightly from that used by some other authors, for example Nadesalingam and Pitman [16]. Similarly, for a system \mathbf{F} as in (1.2), we define

$$\|\mathbf{F}\| = \max_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}} |\lambda_{ij}|.$$

Defining the matrix T as above, we certainly have $\det(T) \neq 0$ and $\|T\| \neq 0$; so we may apply Theorem 1.2 to the system \mathbf{G} with ϵ replaced by the quantity $(|\det(T)|\epsilon) / (R^R \|T\|^{R-1})$. We obtain a nonzero integral solution $\mathbf{x} \in \mathbb{Z}^s$ of the system

$$|G_i(\mathbf{x})| < \frac{|\det(T)|\epsilon}{R^R \|T\|^{R-1}} \quad (1 \leq i \leq R).$$

By Cramer's rule and Hadamard's rule, it follows that \mathbf{x} is also a solution of the system (1.7), whence Theorem 1.1 follows.

We now turn to the proof of Theorem 1.2, which comprises the rest of the paper.

3. INITIAL REDUCTIONS

In this section, we reduce the problem of proving Theorem 1.2 to the consideration of a system of forms as in Theorem 1.2, but under a few more restrictions. This will make our application of the Hardy-Littlewood method easier. We first note that by considering the forms $\epsilon^{-1}F_i$, it is enough to consider only the case

$$\epsilon = 1.$$

We can also assume that we have

$$\|\mathbf{F}\| \geq 1.$$

If this were not the case, then $(1, 0, 0, \dots, 0)$ would be a solution of the system (1.10) and we would be done.

We now quote a lemma, which seems to have been first used in this field by Low, Pitman and Wolff. (See Lemma 1 of [13].) It is actually a special case of a result on matroids, apparently due originally to Edmonds [8]. A proof can also be found in Aigner. (See Proposition 6.45 of [1].)

Lemma 3.1. *Let A be an $R \times s$ matrix over a field K and let w be a positive integer. The matrix A has an $R \times Rw$ partitionable submatrix (that is, A includes w disjoint $R \times R$ submatrices that are nonsingular over K) if and only if the following condition is satisfied:*

$$(3.1) \quad |J| \leq s - w(R - r(A_J)) \quad \text{for all subsets } J \subseteq \{1, 2, \dots, s\}.$$

To be clear, by including w disjoint $R \times R$ nonsingular submatrices, we mean that there is some permutation of the columns so that the first R columns form a nonsingular matrix, as do the second R columns, and so on.

Note that the condition (3.1) is exactly condition (ii) of Theorem 1.2 in the case $w = \ell$. Thus we may apply the lemma to the coefficient matrix A of the system \mathbf{F} , with the choice $w = \ell$. Therefore, A has an $R \times R\ell$ partitionable submatrix. By relabeling variables if necessary, we may write

$$(3.2) \quad A = \begin{bmatrix} A_1 & A_2 & \dots & A_\ell & \lambda_{\ell R+1} & \lambda_{\ell R+2} & \dots & \lambda_s \end{bmatrix},$$

where A_v is an $R \times R$ submatrix for $1 \leq v \leq \ell$ and where

$$(3.3) \quad \Delta_v = |\det(A_v)| \neq 0 \quad \text{for } 1 \leq v \leq \ell.$$

Now consider the system \mathbf{F} in the case that k is odd. We show that \mathbf{F} has a real nonsingular solution. Since A has the form (3.2), and (3.3) holds, one can see that \mathbf{F} is equivalent to a system \mathbf{G} with coefficient matrix B such that the left-hand $R \times 2R$ submatrix of B has the form

$$\begin{bmatrix} I & B_2 \end{bmatrix};$$

here I is the $R \times R$ identity matrix, and B_2 is a nonsingular $R \times R$ matrix. We can find real numbers $z_{R+1}, z_{R+2}, \dots, z_{2R}$ satisfying

$$B_2 \begin{bmatrix} z_{R+1} \\ z_{R+2} \\ \vdots \\ z_{2R} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}.$$

Now let $z_j = 1$ for $1 \leq j \leq R$, and let $z_j = 0$ for $j > 2R$. Then for $1 \leq j \leq s$, define $x_j = z_j^{1/k}$, which is always real, since k is odd. Setting $\mathbf{x} = (x_1, x_2, \dots, x_s)$, one can observe that $G_i(\mathbf{x}) = 0$ for $1 \leq i \leq R$. Now the left-hand $R \times R$ matrix of $\left(\frac{\partial G_i}{\partial x_j} \right)_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}}$ has determinant $k^R x_1^{k-1} x_2^{k-1} \dots x_R^{k-1}$, which is nonzero. Thus the

system \mathbf{G} has a real nonsingular solution, whence, as in Section 2, the system \mathbf{F} does as well.

Thus, whether k is odd or even, we know that there is a nonsingular solution of the system

$$F_i(\mathbf{x}) = 0 \quad (1 \leq i \leq R).$$

We now show that there is a real nonsingular solution whose components are all positive. As noted above, there is a subset $J = \{j_1, j_2, \dots, j_R\} \subseteq \{1, 2, \dots, s\}$ with

$|J| = R$ and a real vector $\mathbf{x} \in \mathbb{R}^s$ such that we have $\det(A_J) \neq 0$ and $\prod_{j \in J} x_j \neq 0$.

Now, for $1 \leq i \leq R$, we define the linear form

$$(3.4) \quad L_i(\mathbf{y}) = L_i(y_1, y_2, \dots, y_s) = \sum_{j=1}^s \lambda_{ij} y_j.$$

On setting $z_j = x_j^k$ for $1 \leq j \leq s$, we see that there is a real vector $\mathbf{z} = (z_1, \dots, z_s) \in \mathbb{R}^s$ such that $z_j \neq 0$ holds for $j \in J$, and we have

$$(3.5) \quad L_i(\mathbf{z}) = 0 \quad \text{for } 1 \leq i \leq R.$$

Now, if k is even, our choice of \mathbf{z} ensures that we have $z_j \geq 0$ for $1 \leq j \leq s$. If k is odd, then for each j , we may if necessary replace z_j by $-z_j$, and replace the coefficients λ_{ij} by $-\lambda_{ij}$ for $1 \leq i \leq R$, and consider the resulting system. In this manner, we may ensure that we have a solution \mathbf{z} of (3.5) with $z_j \geq 0$ for $1 \leq j \leq s$ and $z_j > 0$ for $j \in J$. Note that conditions (ii) and (iii) of Theorem 1.2 and conditions (3.2) and (3.3) are unaffected. Condition (iv) is also unaffected, since the sum $\sum_{x=1}^q e \left(\frac{x^k}{q} \sum_{i=1}^r \lambda_{ij} a_i \right)$ is always real for odd k with $\lambda_{ij} \in \mathbb{Z}$, which may be seen by substituting $-x$ for x in the sum.

Now suppose that $z_{j_0} = 0$ for some j_0 satisfying $1 \leq j_0 \leq s$. We clearly have $j_0 \notin J$. Since we have assumed $\|\mathbf{F}\| \geq 1$, and thus certainly have $\|\mathbf{F}\| \neq 0$, we may fix a positive real number γ with

$$|\gamma| \leq \frac{|\det(A_J)| \min_{j \in J} |z_j|}{2R^R \|\mathbf{F}\|^R}.$$

Since A_J is nonsingular, there is a real vector $\mathbf{w} = (w_1, \dots, w_R)$ such that we have

$$A_J \mathbf{w} = -\gamma \boldsymbol{\lambda}_{j_0}.$$

By Cramer's rule and Hadamard's rule, we certainly have

$$|w_i| \leq \frac{R^R \|\mathbf{F}\|^R \gamma}{|\det(A_J)|} \leq \frac{1}{2} \min_{j \in J} |z_j| \quad \text{for } 1 \leq i \leq R.$$

Now define

$$z'_j = \begin{cases} z_{j_i} + w_i & \text{for } j = j_i \in J \\ \gamma & \text{for } j = j_0 \\ z_j & \text{for } j \notin J \cup \{j_0\}. \end{cases}$$

Writing $\mathbf{z}' = (z'_1, z'_2, \dots, z'_s)$, we have

$$L_i(\mathbf{z}') = L_i(\mathbf{z}) = 0 \quad \text{for } 1 \leq i \leq R.$$

Also, we have $z'_j > 0$ for $j \in J \cup \{j_0\}$. All of the other components of \mathbf{z}' are equal to the respective components of \mathbf{z} ; so we have replaced our real nonsingular solution by a real nonsingular solution that has one more positive component and that still satisfies the condition $z_j > 0$ for $j \in J$. Repeating this process as many as $(s - R)$ times, we can find a nonsingular real solution $\mathbf{z} = (z_1, \dots, z_s)$ with $z_j > 0$ for $1 \leq j \leq s$.

Thus, scaling if necessary, we may choose a real number δ and real numbers z_1, z_2, \dots, z_s that satisfy

$$(3.6) \quad \begin{aligned} 0 < \delta \leq z_j \leq \frac{1}{2} & \quad \text{for } 1 \leq j \leq s, \quad \text{and} \\ L_i(\mathbf{z}) = 0 & \quad \text{for } 1 \leq i \leq R. \end{aligned}$$

To sum up, in this section, we have demonstrated that to prove Theorem 1.2, it is enough to consider a system of forms F_1, F_2, \dots, F_R as in Theorem 1.2, with the added assumptions that $\|\mathbf{F}\| \geq 1$ holds, that the coefficient matrix A of the system satisfies the conditions (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). In sections 4 – 9, we prove Theorem 1.2 under these additional assumptions.

4. THE DAVENPORT-HEILBRONN METHOD: THE SETUP

Now we proceed with the proof of Theorem 1.2 under the additional assumptions we made above. We shall essentially use the Hardy-Littlewood method, in an involved form. We combine the methods of Bentkus and Götze [3] with those of Nadesalingam and Pitman [16].

We note that throughout the paper, implicit constants in the notation $o()$ and $O()$ and \ll and \gg may depend on R, s, k, δ, ℓ , the coefficients of the forms F_1, \dots, F_R , and the real vector \mathbf{z} .

We consider the number of solutions of the system

$$(4.1) \quad \begin{aligned} F_i(\mathbf{y}) &= 0 & \text{for } 1 \leq i \leq r, \\ |F_i(\mathbf{y})| &< 1 & \text{for } r+1 \leq i \leq R. \end{aligned}$$

In the usual fashion, we use a real-valued, even kernel function $K: \mathbb{R} \rightarrow \mathbb{R}$ to give a lower bound for the number of integral solutions of the system (4.1) in a certain range. Define such a function K , for any real number α , by

$$(4.2) \quad K(\alpha) = \left(\frac{\sin \pi \alpha}{\pi \alpha} \right)^2.$$

By Lemma 14.1 of [2], for any real number u , the function K satisfies the identity

$$(4.3) \quad \psi(u) = \int_{-\infty}^{\infty} e(\beta u) K(\beta) d\beta = \begin{cases} 0 & \text{if } |u| \geq 1 \\ 1 - |u| & \text{if } |u| < 1. \end{cases}$$

The function K satisfies, for real numbers β , the bound

$$(4.4) \quad |K(\beta)| \ll \min(1, |\beta|^{-2}).$$

We will also use the identity

$$(4.5) \quad \int_0^1 e(\alpha n) d\alpha = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

Now for positive real numbers P and Q satisfying $Q \leq P$, we define the so-called Q -smooth numbers to be the set

$$\mathcal{A}(P, Q) = \{x \in \mathbb{Z} \text{ with } 1 \leq x \leq P \text{ such that } p|x \implies p \leq Q\}.$$

Fix a positive real number η , to be chosen later, so that it will satisfy the requirements of Lemmas 5.5 and 6.1. Then for real numbers α and P with $P \geq 1$, we define the exponential sum $g(\alpha)$ over the smooth numbers by

$$(4.6) \quad g(\alpha) = g(\alpha, P) = \sum_{x \in \mathcal{A}(P, P^\eta)} e(\alpha x^k).$$

Also, for $1 \leq j \leq s$, and real vectors $\boldsymbol{\alpha} \in \mathbb{R}^R$, we define the linear forms

$$(4.7) \quad \Lambda_j(\boldsymbol{\alpha}) = \sum_{i=1}^R \lambda_{ij} \alpha_i \quad \text{and} \quad \Lambda_j^{(r)}(\boldsymbol{\alpha}) = \sum_{i=1}^r \lambda_{ij} \alpha_i.$$

Then we also define, for $\alpha \in \mathbb{R}^R$ and real numbers P with $P \geq 1$, and for $1 \leq j \leq s$, the functions

$$(4.8) \quad g_j(\alpha) = g_j(\alpha, P) = g(\Lambda_j(\alpha), P).$$

We define as well

$$W = [0, 1]^r \times \mathbb{R}^{R-r}.$$

Now let $\mathcal{N}(P)$ be the number of solutions of the system (4.1) with $x_j \in \mathcal{A}(P, P^\eta)$ for $1 \leq j \leq s$. By using the property (4.3) of the function $K(\alpha)$ and the identity (4.5), one can see that we have

$$\mathcal{N}(P) \geq \sum_{\substack{x_j \in \mathcal{A}(P, P^\eta) \\ 1 \leq j \leq s}} \int_W e \left(\sum_{i=1}^R \alpha_i F_i(\mathbf{x}) \right) \left(\prod_{i=r+1}^R K(\alpha_i) \right) d\alpha;$$

observe that this last remark is justified by the fact that the integral converges absolutely, which follows from (4.4), whence we may write the integral as a product of R integrals. By pulling the sums into the integral, we may rewrite the above bound in the form

$$(4.9) \quad \mathcal{N}(P) \geq \int_W \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha.$$

Thus, to prove Theorem 1.2, it is enough to show that the right-hand side of (4.9) is at least 2.

To this end, we give a dissection of the region of integration W into three subsets. Roughly speaking, we expect that the main contribution to the integral in (4.9) comes from the region where the first r components of α are “close” to rational numbers with small denominators and the last $R - r$ components of α are very small in absolute value. We will show that the contribution to the integral in (4.9) from this region, the so-called major arcs, is positive and “large”, and we will also show that the contribution to the integral from the other regions is smaller, and thus the integral over all of W is positive.

For notational ease, we set

$$(4.10) \quad B = \frac{1}{4(R+1)}.$$

We now define, for positive integers q and integral vectors $\mathbf{a} = (a_1, a_2, \dots, a_r) \in \mathbb{Z}^r$, and real numbers P with $P \geq 2$, the region $\mathcal{M}(q, \mathbf{a})$, or $\mathcal{M}(q, \mathbf{a}, P)$, by

$$(4.11) \quad \begin{aligned} \mathcal{M}(q, \mathbf{a}) = & \left\{ \alpha \in [0, 1]^r \times [-(\log P)^B P^{-k}, (\log P)^B P^{-k}]^{R-r} : \right. \\ & \left. \left\| \alpha_i - \frac{a_i}{q} \right\| \leq (\log P)^B P^{-k} \text{ for } 1 \leq i \leq r \right\}; \end{aligned}$$

here $\|x\|$ denotes the distance from the real number x to the nearest integer. We define the major arcs to be the region

$$(4.12) \quad \mathcal{M} = \mathcal{M}(P) = \bigcup_{1 \leq q \leq (\log P)^B} \bigcup_{\substack{\mathbf{a} \pmod{q} \\ (a_1, \dots, a_r, q) = 1}} \mathcal{M}(q, \mathbf{a}, P),$$

where by $\mathbf{a} \pmod{q}$ we mean that \mathbf{a} runs over vectors $\mathbf{a} \in \mathbb{Z}^r$ such that one has $1 \leq a_i \leq q$ for $1 \leq i \leq r$.

In section 5, we will prove the existence of a function $T(P)$, with $T(P) \geq 1$ and $\lim_{P \rightarrow \infty} T(P) = \infty$, and satisfying a certain property. The function will depend on B and the coefficients of the forms F_i . We define the minor arcs to be the region

$$(4.13) \quad \mathfrak{m} = \mathfrak{m}(P) = ([0, 1]^r \times [-T(P), T(P)]^{R-r}) \setminus \mathcal{M}(P).$$

Finally, we define the trivial arcs to be the set

$$(4.14) \quad \mathfrak{t} = \mathfrak{t}(P) = \{\alpha \in W : |\alpha| > T(P)\}.$$

5. AN ANALOGUE OF WEYL'S INEQUALITY

In this section, we give an analogue of Weyl's inequality. For any real number T with $T \geq 1$, define the region

$$(5.1) \quad \mathfrak{m}_T = \mathfrak{m}_T(P) = ([0, 1]^r \times [-T, T]^{R-r}) \setminus \mathcal{M}(P).$$

We now state the central lemma of this section.

Lemma 5.1. *Fix a positive real number T with $T \geq 1$. Define the forms $F_i(\mathbf{x})$ as in (1.2) for $1 \leq i \leq R$, the region $\mathfrak{m}_T(P)$ as above, and $g_j(\alpha, P)$ for $1 \leq j \leq s$ as in (4.8). Suppose that the coefficient matrix A associated with the system \mathbf{F} has rank R . Suppose also that the irrationality condition (iii) of Theorem 1.2 holds. Then one has*

$$(5.2) \quad \lim_{P \rightarrow \infty} \sup_{\alpha \in \mathfrak{m}_T(P)} \frac{\prod_{j=1}^s |g_j(\alpha, P)|}{P^s} = 0.$$

Observe that trivially one has $\prod_{j=1}^s |g_j(\alpha, P)| \leq P^s$; so we are only seeking a slight improvement over the trivial bound. We also note that the central ideas of the proof stem from the work of Bentkus and Götze [3].

In order to prove Lemma 5.1, we first need to give another lemma, which is essentially a combination of two analogues of Weyl's inequality for exponential sums over smooth numbers. We first quote these two analogues, essentially due to Vaughan and Wooley, as they are presented in [5] as Lemmas 3 and 4, respectively.

Lemma 5.2. *Let α and P be real numbers with $P \geq 2$. Define $g(\alpha) = g(\alpha, P)$ as in (4.6). Fix a positive real number ϵ . Then for sufficiently small η , there is a positive real number γ that depends only on k such that either one has $|g(\alpha, P)| \leq P^{1-\gamma}$, or there are relatively prime integers a and q with $q \geq 1$ that satisfy*

$$g(\alpha, P) \ll q^\epsilon P (q + P^k |q\alpha - a|)^{-1/(2k)} (\log P)^3.$$

Lemma 5.3. *Let α and P be real numbers with $P \geq 3$. Define $g(\alpha) = g(\alpha, P)$ as in (4.6) with $0 < \eta < 1/2$. Fix positive real numbers A and ϵ . Suppose that a and q are relatively prime integers with $1 \leq q \leq (\log P)^A$ and $|q\alpha - a| \leq (\log P)^A P^{-k}$. Then one has*

$$g(\alpha, P) \ll_{A, \epsilon} q^\epsilon P (q + P^k |q\alpha - a|)^{-1/k}.$$

We now state the combination of these lemmas.

Lemma 5.4. Define $\gamma = \gamma(k)$ as in Lemma 5.2. Fix positive real numbers θ and B' . Suppose that P is a real number with $P \geq 3$, and that μ is a real number with

$$(5.3) \quad \mu > \max \left((\log P)^{-B'}, P^{-\gamma} \right).$$

Define $g(\alpha) = g(\alpha, P)$ as in (4.6), with η sufficiently small, and suppose that one has

$$(5.4) \quad |g(\alpha, P)| \geq \mu P.$$

Then there exists a positive integer q and an integer a with $(a, q) = 1$ and

$$q \ll_{B', k, \theta} \mu^{-k-k\theta} \quad \text{and} \quad |q\alpha - a| \ll_{B', k, \theta} \mu^{-k-k\theta} P^{-k}.$$

Proof. It is clearly enough to assume that we have $\theta \leq 1/2$. We apply Lemma 5.2 with the choice $\epsilon = \theta/(2k)$. By (5.3) and (5.4), there exist relatively prime integers a and q with $q \geq 1$ such that one has

$$\mu P \leq |g(\alpha, P)| \ll q^{\theta/(2k)} P (q + P^k |q\alpha - a|)^{-1/(2k)} (\log P)^3.$$

It follows that

$$q^{1-\theta} \ll \mu^{-2k} (\log P)^{6k} \quad \text{and} \quad P^k |q\alpha - a| \ll q^\theta \mu^{-2k} (\log P)^{6k}.$$

By (5.3) and the condition $\theta \leq 1/2$, we certainly have

$$q \ll (\log P)^{\frac{2k(B'+3)}{1-\theta}} \quad \text{and} \quad |q\alpha - a| \ll (\log P)^{\frac{4k(B'+3)}{1-\theta}} P^{-k}.$$

Now we may apply Lemma 5.3, for large P , choosing $A = 5k(B' + 3)/(1 - \theta)$, say, and $\epsilon = \theta/(2k)$. We obtain

$$\mu \ll q^{\theta/(2k)} (q + P^k |q\alpha - a|)^{-1/k}.$$

It follows that one has

$$q \ll \mu^{-k} q^{\theta/2} \quad \text{and} \quad P^k |q\alpha - a| \ll \mu^{-k} q^{\theta/2}.$$

Thus, since $\mu \leq 1$ must hold, we have

$$q \ll \mu^{-\frac{2k}{2-\theta}} \ll \mu^{-k-k\theta} \quad \text{and} \quad |q\alpha - a| \ll \mu^{-k+\frac{\theta}{2}(-k-k\theta)} P^{-k} \ll \mu^{-k-k\theta} P^{-k}.$$

Thus the proof of Lemma 5.4 is complete. \square

Now we are able to give the proof of Lemma 5.1.

Proof. Suppose for the sake of contradiction that the condition (5.2) does not hold. Then there exist a positive real number ϵ , an increasing sequence of positive real numbers P_n with $\lim_{n \rightarrow \infty} P_n = \infty$, and a sequence of real vectors $\alpha_n \in \mathfrak{m}_T(P_n)$ with

$$\prod_{j=1}^s |g_j(\alpha_n, P_n)| > \epsilon P_n^s.$$

We may clearly assume that we have $\epsilon < 1$. By trivial estimates, we have

$$|g_j(\alpha_n, P_n)| > \epsilon P_n \quad \text{for} \quad 1 \leq j \leq s.$$

Now we apply Lemma 5.4 to the sums $g_j(\alpha_n, P_n) = g(\Lambda_j(\alpha_n), P_n)$ for all sufficiently large choices of n . For sufficiently large n , we have the bounds $\epsilon \geq P_n^{-\gamma}$ and

$\epsilon > (\log P_n)^{-1}$, and also $P_n \geq 3$. Thus we may apply Lemma 5.4 with $\mu = \epsilon$ and $\theta = 1/k$ and $B' = 1$. Therefore, there are constants c_1 and c_2 that depend only on k such that for large n and for $1 \leq j \leq s$, there are integers q_{nj} and a_{nj} that satisfy

$$(5.5) \quad 1 \leq q_{nj} \leq c_1 \epsilon^{-k-1} \quad \text{and} \quad |\Lambda_j(\alpha) q_{nj} - a_{nj}| \leq c_2 \epsilon^{-k-1} P_n^{-k}.$$

It follows from these bounds and the definition (5.1) of $\mathbf{m}_T(P_n)$ that we have

$$|a_{nj}| \leq c_2 \epsilon^{-k-1} P_n^{-k} + c_1 \epsilon^{-k-1} RT \|\mathbf{F}\|$$

for all j with $1 \leq j \leq s$, and all large n .

For fixed ϵ and T , we thus have that $|a_{nj}|$ and q_{nj} are uniformly bounded. So there are only finitely many possible $(2s)$ -tuples

$$(q_{n1}, q_{n2}, \dots, q_{ns}, a_{n1}, a_{n2}, \dots, a_{ns}).$$

Therefore one such $(2s)$ -tuple, say $(q_1, \dots, q_s, a_1, \dots, a_s)$, occurs infinitely often. Thus there is some subsequence, say $\{\tilde{n}_m\}$, with

$$(q_{\tilde{n}_m 1}, \dots, q_{\tilde{n}_m s}, a_{\tilde{n}_m 1}, \dots, a_{\tilde{n}_m s}) = (q_1, \dots, q_s, a_1, \dots, a_s)$$

for all $m \in \mathbb{Z}^+$.

Since the sequence $\{\alpha_{\tilde{n}_m}\}$ is contained within the compact set $[0, 1]^r \times [-T, T]^{R-r}$, there is a further subsequence $\{\alpha_{n_m}\}$ and a vector $\alpha_0 \in [0, 1]^r \times [-T, T]^{R-r}$ such that

$$\lim_{m \rightarrow \infty} \alpha_{n_m} = \alpha_0.$$

Our goal in the remainder of the lemma is to show that for sufficiently large values of m , we have $\alpha_{n_m} \in \mathcal{M}(P_{n_m})$, which contradicts our original assumption.

By (5.5) and the defining property of the subsequence $\{\tilde{n}_m\}$, we have

$$(5.6) \quad |\Lambda_j(\alpha_{n_m}) q_j - a_j| \leq c_2 \epsilon^{-k-1} P_{n_m}^{-k} \quad \text{for } 1 \leq j \leq s \quad \text{and for all } m \in \mathbb{Z}^+.$$

Taking the limit of both sides of (5.6) as m goes to infinity, we obtain

$$(5.7) \quad \Lambda_j(\alpha_0) = \frac{a_j}{q_j} \quad \text{for } 1 \leq j \leq s.$$

Because condition (iii) of Theorem 1.2 holds, denoting $\alpha_0 = (\alpha_{01}, \alpha_{02}, \dots, \alpha_{0R})$ we must have

$$(5.8) \quad \alpha_{0(r+1)} = \alpha_{0(r+2)} = \dots = \alpha_{0R}.$$

Therefore we have

$$\Lambda_j^{(r)}(\alpha_0) = \frac{a_j}{q_j} \quad \text{for } 1 \leq j \leq s.$$

Now, by (5.6) and (5.7), we have

$$(5.9) \quad |\Lambda_j(\alpha_{n_m} - \alpha_0)| = \left| \Lambda_j(\alpha_{n_m}) - \frac{a_j}{q_j} + \frac{a_j}{q_j} - \Lambda_j(\alpha_0) \right| \leq c_2 \epsilon^{-k-1} P_{n_m}^{-k}$$

for $1 \leq j \leq s$ and for all $m \in \mathbb{Z}^+$. Now, because A has full rank, we may assume by relabeling variables if necessary that the submatrix A_1 , defined as in (3.2), is nonsingular. Because of this and because the bound (5.9) holds, in particular, for $1 \leq j \leq R$, we must have $|\alpha_{n_m} - \alpha_0| \leq c_3(\mathbf{F}) \epsilon^{-k-1} P_{n_m}^{-k}$ for some constant $c_3 = c_3(\mathbf{F})$ and for all $m \in \mathbb{Z}^+$. Therefore, by (5.8), we must have

$$(5.10) \quad \alpha_{n_m} \in [0, 1]^r \times [-c_3 \epsilon^{-k-1} P_{n_m}^{-k}, c_3 \epsilon^{-k-1} P_{n_m}^{-k}]^{R-r} \quad \text{for } m \in \mathbb{Z}^+.$$

If $r = 0$ holds, then for m sufficiently large, one must have $\alpha_{n_m} \in \mathcal{M}(P_{n_m})$. But this contradicts our original assumption that the sequence α_n satisfies $\alpha_n \in \mathfrak{m}_T(P_n)$, whence the equality (5.2) must hold.

So we may assume for the remainder of the proof that $0 < r \leq R$ holds. Then, using (5.6) and (5.10), for $m \in \mathbb{Z}^+$ and for $1 \leq j \leq s$, we have

$$(5.11) \quad \left| \Lambda_j^{(r)}(\alpha_{n_m}) - \frac{a_j}{q_j} \right| = \left| \Lambda_j(\alpha_{n_m}) - \frac{a_j}{q_j} - \sum_{i=r+1}^R \lambda_{ij} \alpha_{n_m i} \right| \ll \epsilon^{-k-1} P_{n_m}^{-k}.$$

Since A_1 is nonsingular, there is an $r \times r$ submatrix, say A_0 , of A_1 that is nonsingular. We assume for ease of notation that A_0 is the upper left-hand $r \times r$ submatrix of A_1 , noting that the other cases all follow in the same fashion as this case. For any real vector $\alpha = (\alpha_1, \dots, \alpha_R)$, write $\alpha' = (\alpha_1, \dots, \alpha_r)$. By (5.11), we have

$$A_0^T \alpha'_{n_m} = \begin{bmatrix} a_1/q_1 \\ a_2/q_2 \\ \vdots \\ a_r/q_r \end{bmatrix} + \begin{bmatrix} w_{m1} \\ w_{m2} \\ \vdots \\ w_{mr} \end{bmatrix},$$

for some real vector $\mathbf{w}_m = (w_{m1}, \dots, w_{mr})$ with $|\mathbf{w}_m| \ll \epsilon^{-k-1} P_{n_m}^{-k}$. Since we have assumed that A_0 is nonsingular, we may use Cramer's rule to find $\mathbf{b} = (b_1, \dots, b_r)$ with

$$A_0^T \mathbf{b} = \begin{bmatrix} a_1/q_1 \\ a_2/q_2 \\ \vdots \\ a_r/q_r \end{bmatrix}.$$

Since A_0^T has integral entries, one may see that b_i has the form $b_i = d_i/q$ for $1 \leq i \leq r$, where d_i is an integer, and q is a positive integer that satisfies

$$(5.12) \quad q \leq (q_1 q_2 \cdots q_r) \det(A_0^T) \leq c_4(\mathbf{F}) \epsilon^{-r(k+1)},$$

where the last bound follows from (5.5).

We may assume, by reducing if necessary, that we have $(d_1, d_2, \dots, d_r, q) = 1$. By Cramer's rule again, we may find $\mathbf{v}_m \in \mathbb{R}^r$ with $A_0^T \mathbf{v}_m = \mathbf{w}_m$, where we have

$$(5.13) \quad |\mathbf{v}_m| \leq c_5(\mathbf{F}) \epsilon^{-k-1} P_{n_m}^{-k}.$$

Write $\mathbf{d} = (d_1, \dots, d_r)$, and if $d_i = 0$ for some i , define d_i to be q instead. Then we have

$$\alpha'_{n_m} \equiv \frac{1}{q} \mathbf{d} + \mathbf{v}_m \pmod{1} \quad \text{for } m \in \mathbb{Z}^+.$$

Now fix any choice of m large enough so that we have

$$(5.14) \quad (\log P_{n_m})^B \geq \max(c_3(\mathbf{F}), c_4(\mathbf{F}), c_5(\mathbf{F})) \epsilon^{-r(k+1)}.$$

Then by (5.10) we have

$$\alpha_{n_m} \in [0, 1]^r \times [-(\log P_{n_m})^B P_{n_m}^{-k}, (\log P_{n_m})^B P_{n_m}^{-k}]^{R-r}$$

for this choice of m . Now write $\hat{\mathbf{d}} = (d_1, d_2, \dots, d_r, 0, 0, \dots, 0)$, where there are $R-r$ zeros here, and define $\hat{\mathbf{v}}_m$ similarly. Then, setting

$$\mathbf{u} = \hat{\mathbf{v}}_m + (0, 0, \dots, 0, \alpha_{n_m(r+1)}, \dots, \alpha_{n_m R}),$$

we have

$$\alpha_{n_m} \equiv \frac{1}{q} \hat{\mathbf{d}} + \mathbf{u} \pmod{1},$$

where $(d_1, d_2, \dots, d_r, q) = 1$ and where, by combining (5.14) with (5.10), (5.12) and (5.13), we have

$$1 \leq q \leq (\log P_{n_m})^B \quad \text{and} \quad |\mathbf{u}| \leq (\log P_{n_m})^B P_{n_m}^{-k}.$$

Thus, recalling definition (4.11), we have $\alpha_{n_m} \in \mathcal{M}(q, \mathbf{d}, P_{n_m})$ for our particular choice of m and, in fact, recalling (4.12), we also have $\alpha_{n_m} \in \mathcal{M}(P_{n_m})$. As in the case $r = 0$, this is a contradiction, whence the equality (5.2) must in fact hold. This completes the proof of Lemma 5.1. \square

At this point, we make an observation about the lemma for those familiar with earlier arguments of this type. We note that in previous work by the author ([10], [11]), the analogue of our Lemma 5.1 was proved with two different methods, for two subregions of the region $\mathbf{m}_T(P)$. If we were to proceed by analogy with earlier arguments, we would instead have to treat a region $\mathbf{m}_{T, T_0}(P)$, say, in place of $\mathbf{m}_T(P)$, for positive real numbers T_0 with $T_0 \leq T$. The new region would be defined by

$$\mathbf{m}_{T, T_0}(P) = \mathbf{m}_T(P) \cap \{\alpha : |\alpha| \geq T_0\}.$$

Essentially by combining the arguments used for each region in previous proofs, we are able to dispense with the requirement $|\alpha| \geq T_0$.

Having done most of the work, we can now give a lemma that essentially says that $\prod_{j=1}^s |g_j(\alpha, P)|$ is small for $\alpha \in \mathbf{m}$. The idea of using such a lemma is due originally to Bentkus and Götze [3].

Lemma 5.5. *Define the forms $F_i(\mathbf{x})$ as in (1.2) for $1 \leq i \leq R$ and the exponential sums $g_j(\alpha, P)$ for $1 \leq j \leq s$ as in (4.8), with η sufficiently small. Suppose that the coefficient matrix A associated with the system \mathbf{F} has rank R . Suppose also that the irrationality condition (iii) of Theorem 1.2 holds. Then there exists a function $T(P)$ that depends only on B, η and the coefficients of the forms F_1, F_2, \dots, F_R , that satisfies $T(P) \geq 1$ and*

$$(5.15) \quad \lim_{P \rightarrow \infty} T(P) = \infty,$$

and such that if we define $\mathbf{m}(P)$ as in (4.13) with this choice of $T(P)$, then one has

$$\sup_{\alpha \in \mathbf{m}(P)} \prod_{j=1}^s |g_j(\alpha, P)| = o(P^s).$$

Proof. The lemma is very similar to Lemma 6 of [10] and Lemma 4 of [11], and the proof follows in a similar fashion. \square

We note that this lemma (and Lemma 5.1) holds for any positive choice of B , but that the function above that is $o(P^s)$ depends on B . We have stated this lemma in a general fashion in the hopes that it may be useful for future workers.

We observe that one could ensure that the function that is $o(P^s)$ depends only on B, η and $2R - r$ of the coefficients. This follows, with some effort, after finding a subset $J \subseteq \{1, 2, \dots, s\}$ with $|J| = 2R - r$ such that the conditions $\Lambda_j(\alpha) \in \mathbb{Q}$ for $j \in J$, taken together, imply that $\alpha_{r+1} = \dots = \alpha_R = 0$. This can be proved, although our method of proof, at least, is not straightforward.

In the remainder of the paper, we fix a function $T(P)$ that satisfies the conclusions of the above lemma. We note that this is the special function we referred to above in section 4, and is used to define the minor arcs and trivial arcs.

We observe at this point that we could obtain corresponding results which are very similar to Lemmas 5.1 and 5.5 if the exponential sums g_j were replaced by exponential sums over a complete interval. The only major change needed would be to use Lemma 2 of [11] in place of our Lemma 5.4.

6. THE MINOR ARCS

In this section, our goal is to show that the contribution from the minor arcs to the integral in (4.9) is $o(P^{s-Rk})$. We first give a lemma, which is essentially a restatement of results due to Vaughan [19], [20], and results due to Wooley [23].

Lemma 6.1. *Suppose that k is an integer with $k \geq 2$. Define $g(\alpha)$ as in (4.6), with η sufficiently small. Then there is an absolute positive constant C' such that if t is a real number satisfying either*

$$(i) \quad t \geq \min \left(2^k, k(\log k + \log \log k + 2) + \frac{C'k \log \log k}{\log k} \right) \quad \text{for } k \geq 3,$$

or

$$(ii) \quad t > 4 \quad \text{for } k = 2,$$

then one has

$$(6.1) \quad \int_0^1 |g(\alpha)|^t \ll_{\eta} P^{t-k}.$$

We observe that one could certainly improve on the lemma in certain cases, but we choose to use only the above bounds for our results.

Proof. If the first bound of condition (i) holds, then the result is Lemma 6 of [11], which is essentially due to Vaughan [19], [20]. If on the other hand, the second bound of condition (i) holds, then we may essentially quote Lemma 7 of [11], which itself follows almost immediately from work of Wooley [23]. We note that the 3 in Lemma 7 of [11] has been replaced by a 2 here; I am grateful to Scott Parsell for showing me the technique one uses to make this improvement.

In the case in which (ii) holds, we give a proof for completeness. Define

$$\epsilon = t - 4.$$

We need only prove that one has

$$\int_0^1 |g(\alpha)|^{4+\epsilon} d\alpha \ll P^{2+\epsilon}.$$

Clearly, we may assume that $\epsilon \leq 1$ holds. For convenience, we write

$$G = \frac{2}{\epsilon}.$$

Define

$$\mathfrak{N} = \left\{ \alpha \in [0, 1] : |g(\alpha)| > P (\log P)^{-G} \right\}.$$

Also, for positive integers m , define

$$\mathfrak{N}_m = \left\{ \alpha \in \mathfrak{N} : 2^{-m-1}P \leq |g(\alpha)| \leq 2^{-m}P \right\}.$$

Now for $\alpha \in \mathfrak{N}_m$, we apply Lemma 5.4 with the choice $B' = G$. Thus, for large P and any positive real number δ , there exist coprime integers a and q with $q \geq 1$, and

$$q \ll 2^{m(2+\delta)} \quad \text{and} \quad \left| \alpha - \frac{a}{q} \right| \ll q^{-1} 2^{m(2+\delta)} P^{-2}.$$

Thus we have

$$\int_{\mathfrak{N}_m} |g(\alpha)|^{4+\epsilon} d\alpha \ll \sum_{q \ll 2^{m(2+\delta)}} \sum_{a=1}^q (2^{-m} P)^{4+\epsilon} q^{-1} 2^{m(2+\delta)} P^{-2} \ll P^{2+\epsilon} 2^{-m(\epsilon-2\delta)}.$$

It follows for $\delta < \epsilon/2$ that

$$(6.2) \quad \int_{\mathfrak{N}} |g(\alpha)|^{4+\epsilon} d\alpha \ll P^{2+\epsilon} \sum_{m=0}^{\infty} 2^{-m(\epsilon-2\delta)} \ll P^{2+\epsilon}.$$

On the other hand, one has

$$(6.3) \quad \int_{[0,1] \setminus \mathfrak{N}} |g(\alpha)|^{4+\epsilon} d\alpha \ll \left(\sup_{\alpha \in [0,1] \setminus \mathfrak{N}} |g(\alpha)|^{\epsilon} \right) \int_{[0,1]} |g(\alpha)|^4 d\alpha.$$

But $\int_{[0,1]} |g(\alpha)|^4 d\alpha$ is less than or equal to the number of solutions of the equation

$$x_1^2 + x_2^2 = x_3^2 + x_4^2$$

with $1 \leq x_i \leq P$ for $1 \leq i \leq 4$. This is bounded by a constant multiple of $P^2 \log P$, a well-known result, which can be proved by elementary means. Thus from (6.3) and the definition of \mathfrak{N} , we have

$$\int_{[0,1] \setminus \mathfrak{N}} |g(\alpha)|^{4+\epsilon} d\alpha \ll (P(\log P)^{-G})^{\epsilon} P^2 \log P \ll P^{2+\epsilon} (\log P)^{-1},$$

by our choice of G . Combining this bound with (6.2) completes the proof of Lemma 6.1. \square

We note that (6.1) is an example of what one might call an “exact Hua inequality”. In most work using the Hardy-Littlewood method, one uses bounds of the type (6.1) where one only needs to show that for any $\epsilon > 0$, the left side of (6.1) can be bounded by $P^{t-k+\epsilon}$. Dispensing with this ϵ is crucial for our work. The use of such an inequality stems from the work of Bentkus and Götze [3].

For the remainder of the paper, we now fix a choice of η so that Lemmas 5.5 and 6.1 hold for this choice. Now we turn to what is essentially our analogue of Hua’s inequality. It is very similar to Lemma 8 of [10].

Lemma 6.2. *There is an absolute positive real constant \tilde{C}_2 with the following property:*

Assume that the forms F_1, F_2, \dots, F_R are as in Theorem 1.2, with coefficient matrix A satisfying (3.2) and (3.3). Assume that ℓ is a positive integer satisfying

$$\begin{aligned} \ell &\geq 5 && \text{for } k = 2, \text{ and} \\ \ell &\geq \min \left(2^k + 1, k(\log k + \log \log k + 2) + \frac{\tilde{C}_2 k \log \log k}{\log k} \right) && \text{for } k \geq 3. \end{aligned}$$

Define the exponential sums $g_j(\alpha, P)$ as in (4.8), and define the function K as in (4.2). Let $d(P)$ be a nonnegative real-valued function, and let \mathfrak{n} be any subset of the region

$$\mathfrak{d} = \{\alpha \in W \text{ with } |\alpha| \geq d(P)\}.$$

Also, define

$$h(\mathfrak{n}, P) = \sup_{\alpha \in \mathfrak{n}} P^{-s} \prod_{j=1}^s |g_j(\alpha, P)|.$$

Then there is a positive real number ν that depends only on k such that one has

$$\int_{\mathfrak{n}} \prod_{j=1}^s |g_j(\alpha, P)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha \ll (h(\mathfrak{n}, P))^\nu \min(1, d(P)^{-1}) P^{s-Rk}.$$

Proof. Observe first that for any real number ϵ with $0 < \epsilon < 1$, one has

$$\begin{aligned} \int_{\mathfrak{n}} \prod_{j=1}^s |g_j(\alpha, P)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha \\ \ll \left(\sup_{\alpha \in \mathfrak{n}} \prod_{j=1}^s |g_j(\alpha, P)| \right)^\epsilon \int_{\mathfrak{d}} \prod_{j=1}^s |g_j(\alpha, P)|^{1-\epsilon} \prod_{i=r+1}^R |K(\alpha_i)| d\alpha. \end{aligned}$$

It follows from trivial estimates that one has

$$\begin{aligned} \int_{\mathfrak{n}} \prod_{j=1}^s |g_j(\alpha, P)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha \\ \ll (h(\mathfrak{n}, P) P^s)^\epsilon (P^{s-\ell R})^{1-\epsilon} \int_{\mathfrak{d}} \prod_{j=1}^{\ell R} |g_j(\alpha, P)|^{1-\epsilon} \prod_{i=r+1}^R |K(\alpha_i)| d\alpha. \end{aligned}$$

We may certainly choose a positive real number ϵ so that we have

$$(6.4) \quad \ell(1 - \epsilon) > 4 \quad \text{if } k = 2.$$

Defining C' as in Lemma 6.1 and choosing \tilde{C}_2 to be sufficiently large, we may ensure that we have

$$\ell \geq \min \left(2^k + 1, k(\log k + \log \log k + 2) + \frac{C'k \log \log k}{\log k} + 1 \right) \quad \text{if } k \geq 3.$$

Thus we may choose a positive real number ϵ , small enough (in terms only of k and C) so that we have

$$(6.5) \quad \ell(1 - \epsilon) \geq \min \left(2^k, k(\log k + \log \log k + 2) + \frac{C'k \log \log k}{\log k} \right) \quad \text{if } k \geq 3.$$

In each of the cases, we denote our particular choice of ϵ by ν . Now one can join the proof of Lemma 8 of [10] after equation (66), and then follow the remainder of that proof with only slight adjustments. The bounds (6.4) and (6.5) are the crucial bounds that we need to apply Lemma 6.1. We omit the details. \square

Now we can wrap up our work on the minor arcs. We have the following lemma.

Lemma 6.3. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies the conditions (3.2) and (3.3). Choose a function $T(P)$ as in Lemma (5.5). Define the exponential sums $g_j(\alpha)$ as in (4.8), with η sufficiently small, the region \mathfrak{m} as in (4.13), and the function K as in (4.2). Then one has*

$$\int_{\mathfrak{m}} \prod_{j=1}^s |g_j(\alpha)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha = o(P^{s-Rk}).$$

Proof. We simply apply Lemma 6.2 with the choices $\mathfrak{n} = \mathfrak{m}$ and $d(P) = 0$. We have $h(\mathfrak{m}, P) = o(1)$ by Lemma 5.5. Thus the proof of Lemma 6.3 is complete. \square

7. THE TRIVIAL ARCS

In this section we show that the contribution from the trivial arcs to the integral in (4.9) is $o(P^{s-Rk})$, which is now easy to do, having done the necessary work above. We have the following lemma.

Lemma 7.1. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3). Choose a function $T(P)$ as in Lemma (5.5). Define the exponential sums $g_j(\alpha)$ as in (4.8) with η sufficiently small, the region \mathfrak{t} as in (4.14), and the function K as in (4.2). Then one has*

$$\int_{\mathfrak{t}} \prod_{j=1}^s |g_j(\alpha)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha = o(P^{s-Rk}).$$

Proof. We apply Lemma 6.2 with the choices $\mathfrak{n} = \mathfrak{t}$ and $d(P) = T(P)$. We have $h(\mathfrak{m}, P) = O(1)$ by trivial estimates. Thus we obtain

$$\int_{\mathfrak{t}} \prod_{j=1}^s |g_j(\alpha)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha \ll (T(P))^{-1} P^{s-Rk},$$

which by (5.15) of Lemma 5.5 is $o(P^{s-Rk})$. Thus the proof of Lemma 7.1 is complete. \square

8. THE MAJOR ARCS

We now treat the major arcs. Our goal is to show that for large P we have

$$\int_{\mathcal{M}} \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha \gg_{\mathbf{F}} P^{s-Rk}.$$

8.1. Approximation on the major arcs. We start our treatment of the major arcs by approximating the functions $g_j(\alpha)$ by auxiliary functions. We need some notation before we do so. We define Dickman's function ρ by the conditions

$$(8.1) \quad \begin{array}{ll} \rho(u) = 0 & \text{for } u \leq 0, \\ \rho(u) = 1 & \text{for } 0 < u \leq 1, \\ u\rho'(u) = -\rho(u-1) & \text{for } u > 1, \\ \rho \text{ is continuous} & \text{for } u > 0, \\ \rho \text{ is differentiable} & \text{for } u > 1. \end{array}$$

Also, for real numbers β , define the function

$$(8.2) \quad \omega(\beta) = \frac{1}{k} \int_0^{P^k} x^{(1/k)-1} \rho\left(\frac{\log x}{k \log P}\right) e(\beta x) dx.$$

For real vectors $\beta \in \mathbb{R}^R$, define $\Lambda_j(\beta)$ as in (4.7), and write

$$(8.3) \quad \omega_j(\beta) = \omega(\Lambda_j(\beta)) \quad \text{for } 1 \leq j \leq s.$$

Also, for integers q and a with $q \geq 1$, define

$$(8.4) \quad S(q, a) = \sum_{x=1}^q e\left(\frac{ax^k}{q}\right).$$

We now collect some results, given by Brüdern and Cook [5], in the following lemma.

Lemma 8.1. *Define $g(\alpha)$ as in (4.6) and $\omega(\beta)$ as in (8.2). Suppose that a and q are integers with $q \geq 1$, and that β is a real number. Then one has*

$$g\left(\frac{a}{q} + \beta\right) = q^{-1} S(q, a) \omega(\beta) + O\left(\frac{P}{\log P} (q + P^k |\beta|)\right)$$

and

$$\omega(\beta) \ll \min\left(P, |\beta|^{-1/k}\right).$$

Proof. The first result is simply equation (29) of [5]. The second result is essentially the third centered equation on page 135 of [5]. \square

We note that, as remarked by Brüdern and Cook, a and q are not required to be relatively prime. We now state the central lemma of the section, which is very similar to Lemma 4.4 of [16].

Lemma 8.2. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). Define the so-called singular series \mathfrak{S} by*

$$(8.5) \quad \begin{aligned} \mathfrak{S} &= 1 && \text{if } r = 0, \text{ and} \\ \mathfrak{S} &= \sum_{q=1}^{\infty} \sum_{\substack{\mathbf{a} \pmod{q} \\ (a_1, \dots, a_r, q)=1}} q^{-s} \prod_{j=1}^s S\left(q, \Lambda_j^{(r)}(\mathbf{a})\right) && \text{for } 1 \leq r \leq R, \end{aligned}$$

and the singular integral $\mathcal{I}(P)$ by

$$(8.6) \quad \mathcal{I}(P) = \int_{\mathbb{R}^R} \left(\prod_{j=1}^s \omega_j(\beta) \right) \left(\prod_{i=r+1}^R K(\beta_i) \right) d\beta.$$

Fix any positive real number ϵ . Then if P is a sufficiently large positive real number, we have

$$\begin{aligned} \int_{\mathcal{M}} \left(\prod_{j=1}^s g_j(\alpha) \right) \left(\prod_{i=r+1}^R K(\alpha_i) \right) d\alpha - \mathfrak{S} \mathcal{I}(P) \\ \ll P^{s-Rk} \left((\log P)^{2B(R+1)-1} + (\log P)^{B(2-(\ell/k)+\epsilon)} \right). \end{aligned}$$

Proof. There are three steps of the proof. One first approximates each function $g_j(\alpha)$ by terms of the form $q^{-1}S\left(q, \Lambda_j^{(r)}(\mathbf{a})\right)\omega_j(\beta)$ on each of the major arcs, then one extends the integration over each major arc to all of \mathbb{R}^R , and then one extends the sum over q to all positive integers q . The argument closely follows the proof of Lemma 4.4 of [16]. One major difference involves the use of the approximations given in Lemma 8.1. Since we are using exponential sums over smooth numbers, we need to use these approximations instead of more standard results for exponential sums over complete intervals. Finally, we observe that the only condition we really need on ℓ for the purposes of this lemma is the bound $\ell \geq 2k + 1$. \square

Now we turn to consideration of the singular series \mathfrak{S} and the singular integral $\mathcal{I}(P)$. In particular, we shall show that we have $\mathfrak{S} \gg 1$ and also that we have $\mathcal{I}(P) \gg P^{s-Rk}$ for sufficiently large P . We first treat the singular series.

8.2. The singular series. We give some definitions. Suppose that G_1, G_2, \dots, G_r are r integral diagonal forms in s variables with coefficient matrix B , with entries d_{ij} . In Section 8.2, we consider integral forms, and of course we assume that $r \geq 1$ holds throughout this section.

We then define the singular series $\mathfrak{S}(\mathbf{G})$ associated with this system of r forms by

$$(8.7) \quad \mathfrak{S} = \mathfrak{S}(\mathbf{G}) = \sum_{q=1}^{\infty} \sum_{\substack{\mathbf{a} \pmod{q} \\ (a_1, \dots, a_r, q)=1}} q^{-s} \prod_{j=1}^s S(q, M_j(\mathbf{a})),$$

where we set

$$M_j(\mathbf{a}) = \sum_{i=1}^r d_{ij} a_i \quad \text{for } 1 \leq j \leq s.$$

We return to the notation of the rest of the paper for a moment. Observe that the definition of \mathfrak{S} given here coincides with the definition (8.5) in the case $r = R$. Moreover, in general, when $1 \leq r \leq R$ holds, the first singular series is exactly the latter singular series, where the latter is associated with the first r forms F_1, \dots, F_r . Note also that the first singular series is independent of the $R - r$ forms F_{r+1}, \dots, F_R .

Suppose that p is a prime and n is a positive integer. Then we say that an integral vector $\mathbf{x} = (x_1, x_2, \dots, x_s)$ is a **solution of rank $r \pmod{p}$** of the system of congruences

$$(8.8) \quad G_1(\mathbf{x}) \equiv G_2(\mathbf{x}) \equiv \dots \equiv G_r(\mathbf{x}) \equiv 0 \pmod{p^n}$$

if there is a subset $J \subseteq \{1, 2, \dots, s\}$ with $|J| = r$ such that one has $p \nmid \det(B_J)$ and $p \nmid \prod_{j \in J} x_j$. Also, for any prime p and any positive integer n , we define $M(p^n, \mathbf{G})$ to be the number of solutions $\mathbf{x} \pmod{p^n}$ of the system (8.8).

Now we shall define the concept of a normalized system of forms. We follow Low, Pitman and Wolff [13] closely, but we need a slightly more general notion. We essentially want to define a notion of a system such that a related system, which results after setting all but some subset of tr of the variables equal to zero, is normalized in the original sense of Low, Pitman and Wolff.

Suppose that the coefficient matrix B contains t disjoint nonsingular $r \times r$ submatrices B_1, B_2, \dots, B_t . To be clear, by this we mean that there is some permutation of the columns of B so that the first r columns form a nonsingular submatrix, the second r columns form a nonsingular submatrix, and so on, through the t^{th} set of r columns. We define

$$\Delta = \Delta(\mathbf{G}) = \prod_{v=1}^t |\det(B_v)|.$$

Now let $\mathbf{j} = (j_1, j_2, \dots, j_{tr})$ be the ordered (tr) -tuple such that the particular matrix $[B_1 B_2 \dots B_t]$ is the submatrix of B consisting of the columns of B indexed in order by j_1, j_2, \dots, j_{tr} ; that is, we define \mathbf{j} so that for $1 \leq v \leq t$ and $1 \leq h \leq r$, the $j_{v(r-1)+h}^{\text{th}}$ column of B is the h^{th} column of B_v . Observe that the definition of Δ depends on \mathbf{j} . Also set

$$J = \{j_1, j_2, \dots, j_{tr}\}.$$

Suppose that p is a prime dividing Δ . Here we define, following [13] closely, a p -operation on the forms G_1, \dots, G_r as a transformation that produces integral forms H_1, \dots, H_r , and has the following steps:

- (i) Pre-multiply B by an integral unimodular matrix U with entries in the set $\{0, 1, \dots, p-1\}$;
- (ii) Next, multiply at most $tr - r$ of the columns of UB_J by p^k and multiply any of the columns of $UB_{\{1, \dots, s\} \setminus J}$ by p^k ;
- (iii) Then divide g of the rows by p , where we have $1 \leq g \leq r$.

As discussed in [13], step (i) corresponds to adding linear combinations of some of the forms to one or more of the other forms. Step (ii), on the other hand, corresponds to writing $x_j = py_j$ in each column j that one multiplies by p^k , and then trying to solve the new inequalities in the variables y_j . Step (iii) corresponds to dividing g of the r equations by p . One can check, as in [13], that a p -operation is possible for all primes p that divide Δ .

Note that for the resulting system \mathbf{H} , we have

$$\Delta(\mathbf{H}) = p^m \Delta(\mathbf{G})$$

for some integer m . We say that such a p -operation is **permissible** if one has $m < 0$.

Observe that, upon performing permissible p -operations for any of the primes p dividing $\Delta(\mathbf{G})$, we can find a system \mathbf{H} that can be obtained from the original system \mathbf{G} via a finite sequence of permissible p -operations and such that $\Delta(\mathbf{H})$ is minimal. If \mathbf{G} is a system of r integral forms as above, such that $\Delta(\mathbf{G})$ cannot be reduced by any permissible p -operations, then we say that \mathbf{G} is a (\mathbf{j}, t) -**normalized** system. Finally, if \mathbf{G} is a system of r integral forms in exactly tr variables, then we simply say that \mathbf{G} is a **normalized** system. We note that, in this case, our definition clearly agrees with the definition given for a normalized system in [13].

We make one other observation. Suppose that, as above, the coefficient matrix B of a system \mathbf{G} , in s variables with coefficients d_{ij} , contains t disjoint nonsingular $r \times r$ submatrices B_1, B_2, \dots, B_t , and define \mathbf{j} and J as above. Then, we define the system \mathbf{G}^* in tr variables, by defining, for $\mathbf{y} \in \mathbb{Z}^{tr}$ and for $1 \leq i \leq r$, the forms

$$(8.9) \quad G_i^*(\mathbf{y}) = \sum_{n=1}^{tr} d_{ij_n} y_{j_n}^k.$$

Note that this is simply the system obtained by setting all variables with indices $j \notin J$ equal to 0, and subsequently reordering the variables. Observe that the coefficient matrix B^* , say, of the system \mathbf{G}^* has the form

$$B^* = \begin{bmatrix} B_1 & B_2 & \dots & B_t \end{bmatrix}.$$

Suppose now that a system \mathbf{H} can be obtained from the system \mathbf{G} after a finite sequence of permissible p -operations, and suppose that \mathbf{H} is (\mathbf{j}, t) -normalized. Then consider the system \mathbf{H}^* , defined as in (8.9). We can see that the same p -operations (restricted to the columns $j \in J$) allow one to obtain \mathbf{H}^* from the system \mathbf{G}^* ; after all, the variables of \mathbf{G}^* are a subset of those that appear in \mathbf{G} . Each operation is certainly still permissible, since the definition of Δ involves only the columns $j \in J$. If one could reduce $\Delta(\mathbf{H}^*)$ via a permissible p -operation, then we could simply extend step (ii) and multiply all of the columns of $UB_{\{1, \dots, s\} \setminus J}$ by p^k . This would give a permissible p -operation for the system \mathbf{H} , which contradicts our assumption that \mathbf{H} is (\mathbf{j}, t) -normalized. So if \mathbf{H} is (\mathbf{j}, t) -normalized, then it follows that \mathbf{H}^* is normalized, and moreover, if \mathbf{H} results from \mathbf{G} after a finite sequence of permissible p -operations, then \mathbf{H}^* results from \mathbf{G}^* from the same sequence of p -operations.

We can now state the following lemma, which is a step towards bounding the singular series below. We do slightly more than we need to, in the hope that it will be useful for future workers. For this reason, we state the lemma in a self-contained manner.

Lemma 8.3. *Suppose that r , k and s are positive integers with $k \geq 2$, and suppose for $1 \leq i \leq r$ that*

$$D_i(\mathbf{x}) = d_{i1}x_1^k + d_{i2}x_2^k + \dots + d_{is}x_s^k$$

is an integral diagonal form of degree k . Suppose that s satisfies $s \geq tr$, where t is an integer satisfying

$$t \geq 2k + 1.$$

Suppose that the coefficient matrix C of the system \mathbf{D} contains t disjoint nonsingular $r \times r$ submatrices C_1, C_2, \dots, C_t , and define

$$\Delta(\mathbf{D}) = \prod_{v=1}^t |\det(C_v)|.$$

Also, define the singular series $\mathfrak{S}(\mathbf{D})$ as in (8.7). Suppose that the following property, which we denote by $P(t, k, r)$, holds:

Given any system of r integral diagonal forms G_1, G_2, \dots, G_r of degree k in tr variables, with coefficient matrix B which consists of t disjoint nonsingular $r \times r$ submatrices and such that the system \mathbf{G} is normalized, then for every prime p and every positive integer n , there is a solution \mathbf{x} of rank $r \pmod{p}$ of the system of congruences

$$G_1(\mathbf{x}) \equiv G_2(\mathbf{x}) \equiv \dots \equiv G_r(\mathbf{x}) \equiv 0 \pmod{p^n}.$$

Then the series $\mathfrak{S}(\mathbf{D})$ converges absolutely and one has

$$\mathfrak{S}(\mathbf{D}) \gg_{\mathbf{D}} 1.$$

If one also has

$$t \geq kr + k + 1,$$

then there exists a constant $c(k, r, s)$ that depends only on k , r , and s , such that one has

$$\mathfrak{S}(\mathbf{D}) \geq c(k, r, s) (\Delta(\mathbf{D}))^{-3s}.$$

Proof. We first give some more notation. For any prime p , define $\gamma = \gamma(k, p)$ by choosing τ to satisfy $p^\tau \parallel k$, and setting

$$\gamma = \begin{cases} 1 & \text{if } \tau = 0 \\ \tau + 1 & \text{if } \tau > 0 \text{ and } p > 2 \\ \tau + 2 & \text{if } \tau > 0 \text{ and } p = 2. \end{cases}$$

Also, for any prime p , we define

$$(8.10) \quad \chi_{\mathbf{D}}(p) = \lim_{n \rightarrow \infty} \frac{M(p^n, \mathbf{D})}{p^{n(s-r)}}.$$

As in Chapter 5 of [6], and using also Lemma 2.10 of [15], one may see that this limit exists, that $\mathfrak{S}(\mathbf{D})$ converges absolutely, and that $\mathfrak{S}(\mathbf{D})$ is equal to an absolutely convergent product, that is, we have

$$(8.11) \quad \mathfrak{S}(\mathbf{D}) = \prod_p \chi_{\mathbf{D}}(p).$$

(We note that it is in this argument that one uses the condition $t \geq 2k + 1$, and that the rate at which the product converges depends on \mathbf{D} .)

Now define \mathbf{D}^* as in (8.9), and define \mathbf{j} and J as in the discussion above. We may find a (\mathbf{j}, t) -normalized system \mathbf{G} , which can be obtained from \mathbf{D} after a finite sequence of permissible p -operations. As we have noted above, \mathbf{G}^* is then a normalized system, which one obtains from \mathbf{D}^* after (essentially) the same permissible p -operations.

Since property $P(t, k, r)$ holds, there is, for all p and n , a solution \mathbf{w} of rank $r \pmod{p}$ of the system of congruences $\mathbf{G}^*(\mathbf{w}) \equiv \mathbf{0} \pmod{p^n}$. By the way \mathbf{G}^* was defined, one can see that if $\mathbf{y} = (y_1, y_2, \dots, y_s)$ is defined by

$$y_j = \begin{cases} w_j & \text{if } j \in J \\ 0 & \text{if } j \notin J, \end{cases}$$

then we have that \mathbf{y} is a solution of rank $r \pmod{p}$ of the system of congruences $G(\mathbf{y}) \equiv 0 \pmod{p^n}$. In particular, this holds for $n = \gamma$. We may thus apply Lemma 6 of [13] to the system \mathbf{G} , whence we have

$$(8.12) \quad M(p^n, \mathbf{G}) \geq p^{(s-r)(n-\gamma)} \quad \text{for } n > \gamma.$$

From this fact we will deduce a lower bound for $M(p^n, \mathbf{D})$.

To this end, suppose for some positive integer n that $\mathbf{y} \in \mathbb{Z}^s$ is a solution of the congruences

$$G_1(\mathbf{y}) \equiv G_2(\mathbf{y}) \equiv \dots \equiv G_r(\mathbf{y}) \equiv 0 \pmod{p^n}.$$

Recall that the system \mathbf{G} resulted from \mathbf{D} after a finite sequence of permissible q -operations. Let \mathbf{H} be a system such that \mathbf{G} arises from \mathbf{H} after a single permissible q -operation. Let I be the subset of $\{1, 2, \dots, s\}$ consisting of the columns affected by step (ii) of this q -operation, that is, let I consist of the indices such that the

corresponding columns in step (ii) are multiplied by q^k . Then define the vector $\mathbf{x} = (x_1, x_2, \dots, x_s)$ by setting

$$x_j = \begin{cases} qy_j & \text{if } j \in I \\ y_j & \text{if } j \notin I. \end{cases}$$

We show that one has

$$(8.13) \quad H_1(\mathbf{x}) \equiv H_2(\mathbf{x}) \equiv \dots \equiv H_r(\mathbf{x}) \equiv 0 \pmod{p^n}.$$

To see this, let $\mathbf{H}^{(i)}$ and $\mathbf{H}^{(ii)}$ be the systems that result after steps (i) and (ii) of the q -operation, respectively. We certainly have $\mathbf{H}^{(ii)}(\mathbf{y}) \equiv \mathbf{0} \pmod{p^n}$; indeed, some of these forms are congruent to $0 \pmod{p^n q}$. Then observe that we have $\mathbf{H}^{(ii)}(\mathbf{y}) = \mathbf{H}^{(i)}(\mathbf{x})$, whence $\mathbf{H}^{(i)}(\mathbf{x}) \equiv \mathbf{0} \pmod{p^n}$ holds. Since the matrix U is unimodular, so that in particular its determinant is not divisible by p , one has that (8.13) holds. Thus any solution \mathbf{y} of $\mathbf{G}(\mathbf{y}) \equiv \mathbf{0} \pmod{p^n}$ gives rise to a solution \mathbf{x} of $\mathbf{H}(\mathbf{x}) \equiv \mathbf{0} \pmod{p^n}$. If $q \neq p$ holds, we therefore have

$$M(p^n, \mathbf{H}) \geq M(p^n, \mathbf{G}).$$

If $q = p$ holds, we might have some reduction in the number of solutions, because multiplication by p in $\mathbb{Z}/p^n\mathbb{Z}$ has kernel of size p , but we certainly have

$$M(p^n, \mathbf{H}) \geq \frac{M(p^n, \mathbf{G})}{p^{s-r}}.$$

So, by repeating this analysis for each permissible q -operation, one can see that if \mathbf{G} is a (\mathbf{j}, t) -normalized system arising from the system \mathbf{F} after a finite sequence of permissible q -operations, then one has

$$M(p^n, \mathbf{D}) \geq \frac{M(p^n, \mathbf{G})}{p^{m(s-r)}} \quad \text{if } p^m \parallel \Delta(\mathbf{D}).$$

Now the limit $\chi_{\mathbf{G}}(p)$ exists; this follows in much the same way as the corresponding fact for \mathbf{D} . It follows that we have

$$(8.14) \quad \chi_{\mathbf{D}}(p) \geq \frac{\chi_{\mathbf{G}}(p)}{p^{m(s-r)}} \quad \text{if } p^m \parallel \Delta(\mathbf{D}).$$

It follows from (8.12) and the definition of $\chi_{\mathbf{G}}(p)$ that for all primes p we have

$$(8.15) \quad \chi_{\mathbf{D}}(p) \geq p^{-(\gamma + \text{ord}_p(\Delta(\mathbf{D}))(s-r))}.$$

Since the product $\prod_p \chi_{\mathbf{D}}(p)$ is absolutely convergent, there is a constant $c(\mathbf{D})$, which may depend on \mathbf{D} , such that we have

$$\prod_{p > c(\mathbf{D})} \chi_{\mathbf{D}}(p) \geq \frac{1}{2}.$$

Thus, using also (8.15), we have

$$\begin{aligned} \mathfrak{S}(\mathbf{D}) &\geq \frac{1}{2} \prod_{p \leq c(\mathbf{D})} \chi_{\mathbf{D}}(p) \geq \frac{1}{2} \prod_{p \leq c(\mathbf{D})} p^{-(\gamma + \text{ord}_p(\Delta(\mathbf{D}))(s-r))} \\ &\geq \frac{1}{2} (\Delta(\mathbf{D}))^{r-s} \prod_{p \leq c(\mathbf{D})} p^{-\gamma(s-r)} \gg_{\mathbf{D}} 1. \end{aligned}$$

Now suppose that we have $t \geq kr + k + 1$. One can prove as in chapter 5 of [6] that for primes p , one has

$$(8.16) \quad \chi_{\mathbf{D}}(p) = 1 + \sum_{n=1}^{\infty} S(p^n),$$

where for positive integers n , we define

$$S(p^n) = p^{-ns} \sum_{\substack{\mathbf{a} \pmod{p^n} \\ (a_1, \dots, a_r, p)=1}} \prod_{j=1}^s S(p^n, M_j(\mathbf{a})).$$

Now suppose that $p \nmid \Delta(\mathbf{D})$. Suppose for $1 \leq v \leq t$ that C_v consists of the columns $j_{v_1}, j_{v_2}, \dots, j_{v_r}$, in that order. Then for \mathbf{a} satisfying $(a_1, \dots, a_r, p) = 1$, there must exist some $j \in \{j_{v_1}, \dots, j_{v_r}\}$ such that $p \nmid M_j(\mathbf{a})$, since we have $p \nmid \det(C_v)$. Thus for any prime p with $p \nmid \Delta(\mathbf{D})$, and any positive integer n , by the standard estimate $S(p^n, a) \ll p^{n(1-(1/k))}$, which holds for $(a, p) = 1$, we have

$$S(p^n) \ll p^{-ns} \sum_{\substack{\mathbf{a} \pmod{p^n} \\ (a_1, \dots, a_r, p)=1}} p^{ns-(nt/k)} \ll p^{n(r-(t/k))}.$$

Combining this last bound with (8.16) and using $t \geq kr + k + 1$ yields

$$\chi_{\mathbf{D}}(p) - 1 \ll \sum_{n=1}^{\infty} p^{n(r-(t/k))} \ll p^{-1-(1/k)} \quad \text{for } p \nmid \Delta(\mathbf{D}).$$

So there is a constant C that depends only on k, r and t such that one has

$$(8.17) \quad |\chi_{\mathbf{D}}(p) - 1| \leq Cp^{-1-(1/k)} \quad \text{for } p \nmid \Delta(\mathbf{D}).$$

Now, because $\sum_p Cp^{-1-(1/k)}$ converges, there is a constant \tilde{C} that depends only on k, r and t such that one has $1 - Cp^{-1-(1/k)} > 0$ for $p > \tilde{C}$ and

$$\prod_{p > \tilde{C}} \left(1 - Cp^{-1-(1/k)}\right) \geq \frac{1}{2}.$$

Now for all p we have $\chi_{\mathbf{D}}(p) > 0$ from (8.15); so by (8.17) we have

$$\begin{aligned} \prod_p \chi_{\mathbf{D}}(p) &= \prod_{p \leq \tilde{C}} \chi_{\mathbf{D}}(p) \prod_{\substack{p > \tilde{C} \\ p \mid \Delta(\mathbf{D})}} \chi_{\mathbf{D}}(p) \prod_{\substack{p > \tilde{C} \\ p \nmid \Delta(\mathbf{D})}} \chi_{\mathbf{D}}(p) \\ &\geq \prod_{p \leq \tilde{C}} \chi_{\mathbf{D}}(p) \prod_{\substack{p > \tilde{C} \\ p \mid \Delta(\mathbf{D})}} \chi_{\mathbf{D}}(p) \prod_{\substack{p > \tilde{C} \\ p \nmid \Delta(\mathbf{D})}} \left(1 - Cp^{-1-(1/k)}\right) \\ &\geq \frac{1}{2} \prod_{p \leq \tilde{C}} \chi_{\mathbf{D}}(p) \prod_{\substack{p > \tilde{C} \\ p \mid \Delta(\mathbf{D})}} \chi_{\mathbf{D}}(p). \end{aligned}$$

It follows from (8.15) that we have

$$\begin{aligned} \prod_p \chi_{\mathbf{D}}(p) &\geq \frac{1}{2} \prod_{\substack{p \leq \tilde{C} \\ p \nmid \Delta(\mathbf{D})}} p^{-\gamma(s-r)} \prod_{p \mid \Delta(\mathbf{D})} p^{-(\gamma + \text{ord}_p(\Delta(\mathbf{D}))(s-r))} \\ &\geq \frac{1}{2} \left(\prod_{p \leq \tilde{C}} p^{-\gamma(s-r)} \right) (\Delta(\mathbf{D}))^{r-s} \prod_{p \mid \Delta(\mathbf{D})} p^{-\gamma(s-r)} \\ &\gg_{k,r,s} (\Delta(\mathbf{D}))^{r-s} (\Delta(\mathbf{D}))^{2(r-s)}. \end{aligned}$$

This completes the proof of Lemma 8.3. \square

Now we give another lemma that builds on the above lemma and completes the treatment of the singular series for the cases $k \geq 3$. As in the case of Lemma 8.3, we do slightly more than what we will need, and we state the lemma in a self-contained fashion.

Lemma 8.4. *Suppose that r , k , and s are positive integers, and suppose for $1 \leq i \leq r$ that*

$$F_i(\mathbf{x}) = \lambda_{i1}x_1^k + \lambda_{i2}x_2^k + \dots + \lambda_{is}x_s^k$$

is an integral diagonal form of degree k . Suppose that the coefficient matrix A of the system \mathbf{F} contains ℓ disjoint nonsingular $r \times r$ submatrices $A_1^{(r)}, A_2^{(r)}, \dots, A_\ell^{(r)}$, where ℓ is a positive integer satisfying

$$\ell \geq 2k + 1.$$

Define $\Delta = \prod_{v=1}^{\ell} \left| \det \left(A_v^{(r)} \right) \right|$. Define $\mathfrak{S} = \mathfrak{S}(\mathbf{F})$ as in (8.7). Finally, suppose that one of the two following statements holds.

- (i) *k is odd, and $k \geq 3$ holds, and one has $\ell \geq km_0$, where m_0 is the least positive integer m such that one has*

$$2^{m-2} \geq \min \left\{ m^2(2k)^r, (3rk^2)^r \right\}.$$

- (ii) *$k \geq 3$ holds, and one has*

$$\ell > k \left[48k^2 \log 3rk^2 \right].$$

Then one has

$$\mathfrak{S}(\mathbf{F}) \gg_{\mathbf{F}} 1.$$

Moreover, if

$$\ell \geq kr + k + 1$$

holds, then there exists a positive real constant $c(k, r, s)$ that depends only on k, r and s such that one has

$$\mathfrak{S}(\mathbf{F}) \geq c(k, r, s) \Delta^{-3s}.$$

Moreover, we note that if k is odd with $k \geq 3$, then there exists an absolute positive real constant C such that condition (i) holds if one has

$$\ell \geq \frac{rk \log 2k}{\log 2} + Ck \log(r \log 2k).$$

Proof. The last statement of the lemma can be checked with a straightforward computation.

Thus to prove the lemma, we need only check that the condition $P(\ell, k, r)$ of Lemma 8.3 holds for our choices of ℓ . So suppose that \mathbf{G} is a normalized system in ℓr variables. For the case in which condition (i) holds, we may apply Theorem 1(ii) of [13] to see that for any positive integer n , the system $\mathbf{G}(\mathbf{x}) \equiv \mathbf{0} \pmod{p^n}$ has a solution \mathbf{x} of rank $r \pmod{p}$. On the other hand, for the case in which condition (ii) holds, we may apply Theorem 3(i) of [13].

Thus in either case, the condition $P(\ell, k, r)$ of Lemma 8.3 holds, whence Lemma 8.4 follows. □

Now we give a lemma to treat the singular series in the case $k = 2$. We observe that one could surely obtain a result that is better for large r , but we choose not to pursue this here.

Lemma 8.5. *Suppose that r and s are positive integers, and suppose for $1 \leq i \leq r$ that*

$$F_i(\mathbf{x}) = \lambda_{i1}x_1^2 + \lambda_{i2}x_2^2 + \dots + \lambda_{is}x_s^2$$

is an integral diagonal quadratic form. Suppose that the coefficient matrix A , defined as in (1.3), contains ℓ disjoint nonsingular $r \times r$ submatrices, where ℓ is an integer satisfying

$$\ell \geq \min(4r^2 + 4r + 1, 384 \log 16r + 5).$$

Define $\mathfrak{S} = \mathfrak{S}(\mathbf{F})$ as in (8.7). Then one has

$$\mathfrak{S}(\mathbf{F}) \gg_{\mathbf{F}} 1.$$

Proof. If the first bound for ℓ holds, then it follows that any nontrivial complex linear combination of the forms F_1, \dots, F_r has rank at least $4r^2 + 4r + 1$. By the theorem in [18], the singular series \mathfrak{S} is positive and depends only on the forms F_1, F_2, \dots, F_r . One can readily check that the singular series \mathfrak{S} is defined in [18] in the same manner as we have defined it.

Suppose instead that the second bound for ℓ holds. We give a sketch of the proof in this case. We first seek an analogue of Lemma 12 of [13] for the case $k = 2$, with $m \geq [192 \log 16r + 2]$. For primes p with $p \leq 8r^2$, say, it is easy to check that Lemma 5 of [7] provides an analogue of the desired type. To obtain an appropriate analogue of Lemma 12 of [13] for primes $p > 8r^2$, one applies an adaptation of Theorem 2 of [14], with, say, $c = 4$; one can check that if one assumes that the matrix of coefficients contains $c + 1$ nonsingular $r \times r$ submatrices, rather than assuming that the matrix is highly nonsingular, then the result still holds. (This can be seen by noting that the inequality $q_i(\mathbf{B}) > ci$, which would still hold, is the key condition needed on page 339 of [14].)

In either case, we have an analogue of Lemma 12 of [13], and thus we can show that condition $P(\ell, k, r)$ holds, as in the proof of Theorem 1(ii) of [13]. □

Michael Knapp and Professor Wooley provided me with a proof of a result closely related to the second part of the above lemma, for which I am grateful.

8.3. The singular integral. Recall that in (8.6) we defined the singular integral $\mathcal{I}(P)$ by

$$\mathcal{I}(P) = \int_{\mathbb{R}^R} \left(\prod_{j=1}^s \omega_j(\beta) \right) \left(\prod_{i=r+1}^R K(\beta_i) \right) d\beta.$$

Our goal in this section is to demonstrate that for large positive P we have the bound

$$(8.18) \quad \mathcal{I}(P) \gg_{\mathbf{F}} P^{s-Rk}.$$

Instead of using the traditional approach which uses Fourier's Integral Theorem, we use a method given by Schmidt [18]. Below we follow parts of [18] very closely.

Much as in [18], for any positive real number T and any real numbers α and β , we define

$$K_T(\alpha) = K(\alpha T^{-1}) = \left(\frac{\sin(\pi \alpha T^{-1})}{\pi \alpha T^{-1}} \right)^2$$

and

$$(8.19) \quad \psi_T(\beta) = \begin{cases} T(1 - T|\beta|) & \text{for } |\beta| \leq T^{-1} \\ 0 & \text{for } |\beta| > T^{-1}. \end{cases}$$

From (4.3), one may readily deduce the following identity, which holds for all real numbers β , namely,

$$(8.20) \quad \psi_T(\beta) = \int_{\mathbb{R}} e(\alpha\beta) K_T(\alpha) d\alpha.$$

We now define

$$(8.21) \quad \mathcal{I}_T(P) = \int_{\mathbb{R}^R} \left(\prod_{j=1}^s \omega_j(\beta) \right) \left(\prod_{i=1}^r K_T(\beta_i) \right) \left(\prod_{i=r+1}^R K(\beta_i) \right) d\beta.$$

By (4.4) and a similar bound for $K_T(\alpha)$, the integral converges absolutely for each choice of P and T .

We shall see that for fixed P , we have $\lim_{T \rightarrow \infty} \mathcal{I}_T(P) = \mathcal{I}(P)$, and we will also show that for large T , we have $\mathcal{I}_T(P) \gg_{\mathbf{F}} P^{s-Rk}$. These two facts together establish the bound (8.18). To prove the first fact, we give a bound for the difference $\mathcal{I}_T(P) - \mathcal{I}(P)$.

Lemma 8.6. *Suppose that T and P are positive real numbers with $T \geq 1$ and $P \geq 1$. Suppose that R, r, k and s are integers with $R \geq 1$, $0 \leq r \leq R$ and $k \geq 2$. Suppose for $1 \leq i \leq R$ that*

$$F_i(\mathbf{x}) = \lambda_{i1}x_1^2 + \lambda_{i2}x_2^2 + \dots + \lambda_{is}x_s^2$$

is a real diagonal form of degree k and that for $1 \leq i \leq r$, the form F_i is integral. Assume that one has $\|\mathbf{F}\| \geq 1$. Suppose also that the coefficient matrix A of the system \mathbf{F} is as in (3.2) and satisfies (3.3), where one has

$$\ell \geq k + 1.$$

Define $\mathcal{I}(P)$ and $\mathcal{I}_T(P)$ as in (8.6) and (8.21), respectively. Then one has

$$\mathcal{I}(P) - \mathcal{I}_T(P) \ll T^{-1/k} P^{s-Rk} \left(\|\mathbf{F}\|^{2R} \prod_{v=1}^{\ell} \Delta_v^{-3/\ell} + P^{k-\ell} \|\mathbf{F}\|^{(R\ell)/k} \prod_{v=1}^{\ell} \Delta_v^{-1/k} \right).$$

We note that the implicit constant in Vinogradov’s notation here depends at most on R, r, k and s and, in particular, does not depend on the coefficients of \mathbf{F} .

Proof. Observe first that, in the case $r = 0$, we have $\mathcal{I}(P) = \mathcal{I}_T(P)$. So we can assume that we have $r \geq 1$. It follows from the definitions (8.6) and (8.21) that we have

$$\mathcal{I}(P) - \mathcal{I}_T(P) \ll \int_{\mathbb{R}^R} \left(\prod_{j=1}^s |\omega_j(\boldsymbol{\beta})| \right) \left| 1 - \prod_{i=1}^r K_T(\beta_i) \right| \left(\prod_{i=r+1}^R |K(\beta_i)| \right) d\boldsymbol{\beta}.$$

From the penultimate centered equation on page 305 of [18], one has

$$1 - \prod_{i=1}^r K_T(\beta_i) \ll T^{-2} \max_{1 \leq i \leq r} |\beta_i|^2 \ll T^{-2} |\boldsymbol{\beta}|^2 \quad \text{for } |\boldsymbol{\beta}| < T,$$

and for $|\boldsymbol{\beta}| \geq T$, one clearly has $\prod_{i=1}^r K_T(\beta_i) \ll 1$. Combining these bounds with the estimate for $\omega(\boldsymbol{\beta})$ given in Lemma 8.1, and the bound (4.4) for $K(\alpha)$, one has

(8.22)

$$\begin{aligned} \mathcal{I}(P) - \mathcal{I}_T(P) \ll & \quad T^{-2} P^{s-R\ell} \int_{|\boldsymbol{\beta}| < T} \left(\prod_{j=1}^{R\ell} \min \left(P, |\Lambda_j(\boldsymbol{\beta})|^{-1/k} \right) \right) |\boldsymbol{\beta}|^2 d\boldsymbol{\beta} \\ & + P^{s-R\ell} \int_{|\boldsymbol{\beta}| \geq T} \prod_{j=1}^{R\ell} \min \left(P, |\Lambda_j(\boldsymbol{\beta})|^{-1/k} \right) d\boldsymbol{\beta}. \end{aligned}$$

Consider the first integral on the right-hand side of (8.22). By Hölder’s inequality, one has

(8.23)

$$\begin{aligned} & \int_{|\boldsymbol{\beta}| < T} |\boldsymbol{\beta}|^2 \prod_{j=1}^{R\ell} \min \left(P, |\Lambda_j(\boldsymbol{\beta})|^{-1/k} \right) d\boldsymbol{\beta} \\ & \ll \prod_{v=1}^{\ell} \left(\int_{|\boldsymbol{\beta}| < T} |\boldsymbol{\beta}|^2 \prod_{j=(v-1)R+1}^{vR} \min \left(P, |\Lambda_j(\boldsymbol{\beta})|^{-1/k} \right)^{\ell} d\boldsymbol{\beta} \right)^{1/\ell}. \end{aligned}$$

For a fixed choice of v with $1 \leq v \leq \ell$, one makes the change of variable $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_R) = U_v(\boldsymbol{\beta})$ given by $\gamma_j = \Lambda_{(v-1)R+j}(\boldsymbol{\beta})$ for $1 \leq j \leq R$, and obtains

$$\begin{aligned} & \int_{|\boldsymbol{\beta}| < T} |\boldsymbol{\beta}|^2 \prod_{j=(v-1)R+1}^{vR} \min \left(P, |\Lambda_j(\boldsymbol{\beta})|^{-1/k} \right)^{\ell} d\boldsymbol{\beta} \\ & \ll \Delta_v^{-1} \int_{\boldsymbol{\gamma} \in U_v([-T, T]^R)} |U_v^{-1}(\boldsymbol{\gamma})|^2 \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^{\ell} d\boldsymbol{\gamma}. \end{aligned}$$

If $\boldsymbol{\gamma} = U_v(\boldsymbol{\beta})$, then one has

$$|\boldsymbol{\gamma}| = |U_v(\boldsymbol{\beta})| \leq R\|\mathbf{F}\| \cdot |\boldsymbol{\beta}|,$$

and by Cramer's rule and Hadamard's inequality, one has

$$(8.24) \quad |\beta| = |U_v^{-1}(\gamma)| \leq \frac{R^{R/2} \|\mathbf{F}\|^{R-1}}{\Delta_v} |\gamma|.$$

It follows that we have

$$(8.25) \quad \int_{|\beta| < T} |\beta|^2 \prod_{j=(v-1)R+1}^{vR} \min \left(P, |\Lambda_j(\beta)|^{-1/k} \right)^\ell d\beta \\ \ll_R \Delta_v^{-3} \|\mathbf{F}\|^{2R-2} \int_{|\gamma| \leq R \|\mathbf{F}\| T} |\gamma|^2 \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^\ell d\gamma.$$

But one has

$$\int_{|\gamma| \leq R \|\mathbf{F}\| T} |\gamma|^2 \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^\ell d\gamma \\ \ll \sum_{n=1}^{\lceil R \|\mathbf{F}\| T \rceil} n^2 \int_{n-1 \leq |\gamma| < n} \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^\ell d\gamma \\ \ll \sum_{n=1}^{\lceil R \|\mathbf{F}\| T \rceil} n^2 \left(\int_0^n \min \left(P, \gamma^{-1/k} \right)^\ell d\gamma \right)^{R-1} \int_{n-1}^n \min \left(P, \gamma^{-1/k} \right)^\ell d\gamma \\ \ll \sum_{n=1}^{\lceil R \|\mathbf{F}\| T \rceil} n^{2P^{(\ell-k)(R-1)}} \min \left(P^{\ell-k}, (n-1)^{-\frac{\ell}{k}} \right),$$

whence we have

$$\int_{|\gamma| \leq R \|\mathbf{F}\| T} |\gamma|^2 \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^\ell d\gamma \\ \ll P^{(\ell-k)(R-1)} \left(P^{\ell-k} + \sum_{n=1}^{\lceil R \|\mathbf{F}\| T \rceil} n^{2-(\ell/k)} \right) \\ \ll P^{R(\ell-k)} + P^{(\ell-k)(R-1)} \sum_{n=1}^{\lceil R \|\mathbf{F}\| T \rceil} n^{1-(1/k)} \\ \ll P^{R(\ell-k)} (\|\mathbf{F}\| T)^{2-(1/k)},$$

since we have $\ell \geq k+1$ and $\|\mathbf{F}\| T \geq 1$.

Thus, by (8.22), (8.23) and (8.25), we have

$$\mathcal{I}(P) - \mathcal{I}_T(P) \ll \|\mathbf{F}\|^{2R} T^{-1/k} P^{s-Rk} \prod_{v=1}^{\ell} \Delta_v^{-3/\ell} \\ + P^{s-R\ell} \int_{|\beta| \geq T} \prod_{j=1}^{R\ell} \min \left(P, |\Lambda_j(\beta)|^{-1/k} \right) d\beta.$$

By Hölder's inequality and a change of variable as above, we have

$$\begin{aligned} \mathcal{I}(P) - \mathcal{I}_T(P) &\ll \| \mathbf{F} \|^{2R} T^{-1/k} P^{s-Rk} \prod_{v=1}^{\ell} \Delta_v^{-3/\ell} \\ &\quad + P^{s-R\ell} \prod_{v=1}^{\ell} \left(\Delta_v^{-1} \int_{|\gamma| \geq c_{\mathbf{F},v} T} \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^{\ell} d\gamma \right)^{1/\ell}, \end{aligned}$$

where $c_{\mathbf{F},v}$ is a positive constant which by (8.24) we may define by

$$c_{\mathbf{F},v} = \frac{\Delta_v}{R^{R/2} \| \mathbf{F} \|^{R-1}}.$$

It follows that one has

$$\begin{aligned} &\int_{|\gamma| \geq c_{\mathbf{F},v} T} \prod_{j=1}^R \min \left(P, |\gamma_j|^{-1/k} \right)^{\ell} d\gamma \\ &\ll \left(\int_0^{\infty} \min \left(P, |\gamma|^{-1/k} \right)^{\ell} d\gamma \right)^{R-1} \int_{c_{\mathbf{F},v} T}^{\infty} \min \left(P, |\gamma|^{-1/k} \right)^{\ell} d\gamma \\ &\ll P^{(\ell-k)(R-1)} (c_{\mathbf{F},v} T)^{1-(\ell/k)} \\ &\ll P^{(\ell-k)R-\ell+k} \Delta_v^{1-(\ell/k)} \| \mathbf{F} \|^{(R\ell)/k} T^{-1/k}, \end{aligned}$$

since we have $\ell \geq k+1$. Combining the last two bounds completes the proof of Lemma 8.6. \square

Now we prove a lemma which states that for T and P sufficiently large, the quantity $\mathcal{I}_T(P)$ is bounded below.

Lemma 8.7. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). Suppose also that one has $\| \mathbf{F} \| \geq 1$. Define Δ_1 as in (3.3). Suppose that T and P are real numbers satisfying $T \geq 1$ and*

$$(8.26) \quad P \geq \left(\frac{12}{\delta} s^R R^{2R} \| \mathbf{F} \|^{2R} \max(\Delta_1^{-2}, 1) \right)^{1/k},$$

where δ is as in (3.6). Define $\mathcal{I}_T(P)$ as in (8.21). Then there is a constant $c_1 = c_1(\mathbf{F}, k, R, r, s, \delta, \mathbf{z})$ that does not depend on T or P such that one has

$$\mathcal{I}_T(P) \geq c_1 P^{s-Rk}.$$

Proof. Recalling the definition (8.3) of $\omega_j(\beta)$, we can write the absolutely convergent integral $\mathcal{I}_T(P)$ in the form

$$\begin{aligned} &\frac{1}{k^s} \int_{\mathbb{R}^R} \int_{[0, P^k]^s} (x_1 x_2 \cdots x_s)^{1/k-1} \left(\prod_{j=1}^s \rho \left(\frac{\log x_j}{k \log P} \right) \right) e \left(\sum_{j=1}^s \Lambda_j(\beta) x_j \right) \\ &\quad \times \prod_{i=1}^r K_T(\beta_i) \prod_{i=r+1}^R K(\beta_i) d\mathbf{x} d\beta. \end{aligned}$$

Using the identities (8.20) and (4.3), and recalling the definition (3.4) of $L_i(\mathbf{x})$, we can rewrite $\mathcal{I}_T(P)$ as

$$\frac{1}{k^s} \int_{[0, P^k]^s} (x_1 \cdots x_s)^{1/k-1} \prod_{j=1}^s \rho\left(\frac{\log x_j}{k \log P}\right) \left(\prod_{i=1}^r \psi_T(L_i(\mathbf{x}))\right) \left(\prod_{i=r+1}^R \psi(L_i(\mathbf{x}))\right) d\mathbf{x}.$$

$\mathcal{I}_T(P)$ is certainly larger than the corresponding integral over the smaller region $\left[\frac{\delta P^k}{2}, P^k\right]^s$, noting from Lemma 12.1(i) of [21], and (8.1), that the function ρ is always nonnegative. For P satisfying (8.26), one certainly has $P > (\delta/2)^{-1/k}$, whence for x_j satisfying $\frac{\delta P^k}{2} \leq x_j \leq P^k$, it follows from (8.1) that one has

$$\rho\left(\frac{\log x_j}{k \log P}\right) = 1.$$

So we have

$$\mathcal{I}_T(P) \geq \frac{1}{k^s} \int_{\left[\frac{\delta P^k}{2}, P^k\right]^s} (x_1 x_2 \cdots x_s)^{1/k-1} \left(\prod_{i=1}^r \psi_T(L_i(\mathbf{x}))\right) \left(\prod_{i=r+1}^R \psi(L_i(\mathbf{x}))\right) d\mathbf{x}.$$

Now define $\mathcal{R}_{P,T}$ to be the region

$$\left\{ \mathbf{x} \in \left[\frac{\delta P^k}{2}, P^k\right]^s : |L_i(\mathbf{x})| < \frac{\min(1, \Delta_1^{1/R})}{3T} \text{ for } 1 \leq i \leq r \text{ and } |L_i(\mathbf{x})| < \frac{1}{3} \min(1, \Delta_1^{1/R}) \text{ for } r+1 \leq i \leq R \right\}.$$

It follows from the definitions (8.19) and (4.3) of $\psi_T(\alpha)$ and $\psi(\alpha)$, respectively, that we have

$$(8.27) \quad \mathcal{I}_T(P) \gg P^{s-sk} T^r \mu_s(\mathcal{R}_{P,T}),$$

where μ_s denotes s -dimensional Euclidean measure.

We now make the linear change of variable $\mathbf{w} = V(\mathbf{x})$ given by

$$w_j = \begin{cases} L_j(\mathbf{x}) & \text{for } 1 \leq j \leq R \\ x_j & \text{for } R+1 \leq j \leq s. \end{cases}$$

Since Δ_1 is nonzero, we can see that we have

$$(8.28) \quad \mu_s(\mathcal{R}_{P,T}) \gg_{\mathbf{F}} \mu_s(\mathcal{S}_{P,T}),$$

where we define $\mathcal{S}_{P,T}$ to be the region $V(\mathcal{R}_{P,T})$. Note that $\mathcal{S}_{P,T}$ is the set of $\mathbf{w} \in \mathbb{R}^s$ such that there exists an $\mathbf{x} \in \left[\frac{\delta P^k}{2}, P^k\right]^s$ with $\mathbf{w} = V(\mathbf{x})$ and such that one has

$$|w_i| < \frac{\min(1, \Delta_1^{1/R})}{3T} \text{ for } 1 \leq i \leq r \text{ and } |w_i| < \frac{\min(1, \Delta_1^{1/R})}{3} \text{ for } r+1 \leq i \leq R.$$

Now we give a lemma, which is essentially due to Nadesalingam and Pitman. (See [16], Lemma 5.2.)

Lemma 8.8. *Let R and s be positive integers satisfying $s > R$. Let*

$$A = (\lambda_{ij})_{\substack{1 \leq i \leq R \\ 1 \leq j \leq s}}$$

be a real $R \times s$ matrix. For $1 \leq i \leq R$, we define the linear forms $L_i(\mathbf{y}) = \sum_{j=1}^s \lambda_{ij} y_j$.

Let Δ_1 denote the absolute value of the determinant of the left-hand $R \times R$ submatrix of A . Suppose that we have $\Delta_1 > 0$. Additionally, suppose that Q is a real number satisfying

$$(8.29) \quad Q \geq 6s^R R^{2R} \|A\|^{2R} \Delta_1^{-2}.$$

Suppose also that w_1, \dots, w_R are real numbers satisfying

$$|w_i| \leq \frac{\Delta_1^{1/R}}{3} \quad \text{for } 1 \leq i \leq R.$$

Let $S_0 = S_0(w_1, \dots, w_R)$ be the set of all real vectors $(y_{R+1}, \dots, y_s) \in [-Q, Q]^{s-R}$ for which there exist real numbers $y_1, \dots, y_R \in [-Q, Q]$ with $w_i = L_i(\mathbf{y})$ for $1 \leq i \leq R$.

Then S_0 has $(s - R)$ -dimensional measure satisfying

$$\mu_{s-R}(S_0) \gg_A Q^{s-R},$$

where the implicit constant in Vinogradov's notation depends on s and R and the entries of A .

Proof. We apply the lemma of Nadesalingam and Pitman to the R linear forms $M_1(\mathbf{y}), \dots, M_R(\mathbf{y})$ defined by $M_i(\mathbf{y}) = \Delta_1^{-1/R} L_i(\mathbf{y})$ for $1 \leq i \leq R$, in order to relax the requirement $\Delta_1 \geq 1$ of their lemma. We note that there is a slight difference between the definition of $\|A\|$ that we use and the definition they use, which accounts for the change in the condition (8.29). Here we have also implicitly used the last equation on page 704 of [15] to show that the term $H(L)$ in the lemma of Nadesalingam and Pitman is positive. \square

Now we return to the proof of Lemma 8.7 and apply Lemma 8.8. By (8.26) and the assumption $T \geq 1$, we may apply the lemma, with the choice $Q = (\delta P^k)/2$, for any $\mathbf{w} = (w_1, w_2, \dots, w_R)$ with

$$(8.30) \quad |w_i| < \frac{\min(1, \Delta_1^{1/R})}{3T} \text{ for } 1 \leq i \leq r, \quad \text{and} \quad |w_i| < \frac{\min(1, \Delta_1^{1/R})}{3} \text{ for } r+1 \leq i \leq R.$$

We obtain

$$\mu_{s-R}(S_0(w_1, \dots, w_R)) \gg_{\mathbf{F}, \delta} P^{k(s-R)}.$$

Now, for any choice of $(y_{R+1}, y_{R+2}, \dots, y_s) \in S_0(w_1, \dots, w_R)$, there exist real numbers $y_1, \dots, y_R \in [-Q, Q]$ with $L_i(\mathbf{y}) = w_i$ for $1 \leq i \leq R$. Defining \mathbf{z} as in (3.6), we have

$$L_i(P^k \mathbf{z} + \mathbf{y}) = w_i \quad \text{for } 1 \leq i \leq R.$$

By (3.6) and our choice of Q , we also have

$$P^k \mathbf{z} + \mathbf{y} \in \left[\frac{\delta}{2} P^k, \frac{1+\delta}{2} P^k \right] \subseteq \left[\frac{\delta}{2} P^k, P^k \right].$$

Recalling the definition of $\mathcal{S}_{P,T}$, we see that we have

$$\mu_s(\mathcal{S}_{P,T}) \gg_{\mathbf{F},\delta} T^{-r} P^{k(s-R)}.$$

Combining with (8.27) and (8.28), we see that there is a positive real constant $c_1 = c_1(\mathbf{F}, k, R, r, s, \delta, \mathbf{z})$ such that one has

$$(8.31) \quad \mathcal{I}_T(P) \geq c_1 P^{s-Rk}.$$

This completes the proof of Lemma 8.7. \square

Combining Lemmas 8.6 and 8.7 yields the following lower bound for $\mathcal{I}(P)$.

Lemma 8.9. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). Assume also that we have $\|\mathbf{F}\| \geq 1$. Define $\mathcal{I}(P)$ as in (8.6). Then there is a constant $c_3 = c_3(\mathbf{F}, k, R, r, s, \delta, \mathbf{z})$ such that for $P \geq c_3$, one has*

$$\mathcal{I}(P) \gg P^{s-Rk}.$$

Here the implicit constant in Vinogradov's notation may depend on $\mathbf{F}, k, s, R, r, \delta$ and the special real vector \mathbf{z} , but it does not depend on P .

8.4. Completion of the treatment of the major arcs. We wrap up our work on the major arcs with the following lemma.

Lemma 8.10. *Suppose that we are in the setting of Theorem 1.2 and that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). Assume also that we have $\|\mathbf{F}\| \geq 1$. Then there are constants c_4 and c_5 , which may depend on $\mathbf{F}, k, s, R, r, \delta$ and \mathbf{z} , but which do not depend on P , such that for real numbers P satisfying $P \geq c_4$, one has*

$$\int_{\mathcal{M}} \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha \geq c_5 P^{s-Rk}.$$

Proof. Choose $\epsilon = 1/(2k)$ and apply Lemma 8.2. Since we have $\ell \geq 2k + 1$ and by the definition (4.10) of B , we obtain

$$(8.32) \quad \int_{\mathcal{M}} \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha - \mathfrak{S} \mathcal{I}(P) \ll P^{s-Rk} (\log P)^{-1/(8k(R+1))}.$$

Since condition (iv) of Theorem 1.2 holds, one has $\mathfrak{S} \gg 1$. By Lemma 8.9, there are constants c_3 and c_6 that do not depend on P such that one has

$$(8.33) \quad \mathcal{I}(P) \geq c_6 P^{s-Rk} \quad \text{for } P \geq c_3.$$

Lemma 8.10 follows from (8.32) and (8.33) and the bound $\mathfrak{S} \gg 1$. \square

9. COMPLETION OF THE PROOF OF THEOREM 1.2

In this section, we gather together all of our results in order to complete the proof of Theorem 1.2.

We recall that we demonstrated in Section 3 that we may assume that we have $\epsilon = 1$, that we have $\|\mathbf{F}\| \geq 1$, that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6).

We first observe how one proves the last sentence of Theorem 1.2, namely that if we have $r \geq 1$, and we define $m_0(r, k, \tilde{C}_1)$ as in (1.4) and assume that we have $\ell \geq m_0(r, k, \tilde{C}_1)$, then we have $\mathfrak{S} \gg_{\mathbf{F}} 1$. For $k \geq 3$, one simply applies Lemma 8.4, noting that we certainly have $\ell \geq 2k + 1$ for a sufficiently large choice of the constant \tilde{C}_1 . For $k = 2$, we may apply Lemma 8.5. Thus we have $\mathfrak{S} \gg_{\mathbf{F}} 1$.

Now we turn to the central result of Theorem 1.2. Recall from (4.9) that we have

$$(9.1) \qquad \mathcal{N}(P) \geq \int_W \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha,$$

where $\mathcal{N}(P)$ was defined to be the number of solutions of the system (4.1) with $x_j \in \mathcal{A}(P, P^\eta)$ for $1 \leq j \leq s$.

We first choose a function $T(P)$ as in Lemma 5.5. We can now treat the minor arcs and trivial arcs. By Lemmas 6.3 and 7.1, one obtains

$$(9.2) \qquad \int_{\mathfrak{m} \cup \mathfrak{t}} \prod_{j=1}^s |g_j(\alpha)| \prod_{i=r+1}^R |K(\alpha_i)| d\alpha = o(P^{s-Rk}).$$

We now consider the major arcs. By Lemma 8.10, we have

$$\int_{\mathcal{M}} \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha \geq c_5 P^{s-Rk}$$

for $P \geq c_4$, where c_4 and c_5 are constants that do not depend on P . Together with (9.2), it follows for sufficiently large P that one has

$$\int_W \prod_{j=1}^s g_j(\alpha) \prod_{i=r+1}^R K(\alpha_i) d\alpha \geq \frac{c_5}{2} P^{s-Rk}.$$

By (9.1), for sufficiently large P , we have

$$\mathcal{N}(P) \gg_{\mathbf{F}, R, s, k} P^{s-Rk}.$$

This establishes Theorem 1.2.

As a final observation, we note that we have obtained a lower bound of the expected order of magnitude for the number of solutions of our system in a box of size P , for all sufficiently large positive P . Recall that we assumed that we have $\epsilon = 1$, that we have $\|\mathbf{F}\| \geq 1$, that the coefficient matrix A of the system \mathbf{F} satisfies (3.2) and (3.3), and that there is a real vector \mathbf{z} satisfying (3.6). Using standard techniques, one can check that under the conditions of either Theorem 1.1 or Theorem 1.2, without any of these simplifying assumptions, the same lower bound holds for sufficiently large P . We note that in this case, P must be sufficiently large also in terms of ϵ , and the implicit constant in the lower bound for $\mathcal{N}(P)$ depends on ϵ .

REFERENCES

1. M. Aigner, *Combinatorial theory*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, New York/Heidelberg/Berlin, 1979. MR 80h:05002
2. R. C. Baker, *Diophantine inequalities*, London Mathematical Society Monographs, New Series, 1, The Clarendon Press, Oxford University Press, New York, 1986. MR 88f:11021

3. V. Bentkus and F. Götze, *Lattice point problems and distribution of values of quadratic forms*, Ann. of Math. (2) **150** (1999), no. 3, 977–1027. MR **2001b**:11087
4. B. J. Birch and H. Davenport, *Indefinite quadratic forms in many variables*, Mathematika, **5** (1958), 8–12. MR **20**:3104
5. J. Brüdern and R. J. Cook, *On simultaneous diagonal equations and inequalities*, Acta Arith. **62** (1992), 125–149. MR **93h**:11036
6. H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities*, Ann Arbor Publishers, Ann Arbor, MI, 1963. MR **28**:3002
7. H. Davenport and D. J. Lewis, *Simultaneous equations of additive type*, Philos. Trans. Roy. Soc. London Ser. A **264** (1969), 557–595. MR **39**:6848
8. J. Edmonds, *Minimum partition of a matroid into independent subsets*, J. Res. Nat. Bureau Standards **69B** (1965), 67–72. MR **32**:7441
9. D. E. Freeman, *A note on one cubic Diophantine inequality*, J. London Math. Soc. (2), **61** (2000), no.1, 25–35. MR **2001c**:11043
10. ———, *Quadratic Diophantine inequalities*, J. Number Theory, **89** (2001), no.2, 268–307.
11. ———, *Asymptotic lower bounds for Diophantine inequalities*, to appear in Mathematika.
12. ———, *Asymptotic lower bounds and formulas for Diophantine inequalities*, to appear in the Proceedings of the Millennial Conference in Number Theory.
13. L. Low, J. Pitman and A. Wolff, *Simultaneous diagonal congruences*, J. Number Theory **29** (1988), 31–59. MR **89g**:11030
14. I. D. Meir, *Simultaneous diagonal p -adic equations*, Mathematika **45** (1998), 337–349. MR **2000k**:11052
15. T. Nadesalingam and J. Pitman, *Bounds for solutions of simultaneous diagonal equations of odd degree*, Théorie des nombres (Québec, PQ, 1987), 703–734, de Gruyter, Berlin, 1989. MR **91f**:11021
16. ———, *Simultaneous diagonal inequalities of odd degree*, J. Reine Angew. Math., **394** (1989), 118–158. MR **91c**:11019
17. W. M. Schmidt, *Diophantine inequalities for forms of odd degree*, Adv. Math. **38** (1980), 128–151. MR **82h**:10033
18. ———, *Simultaneous rational zeros of quadratic forms*, Seminar Delange-Pisot-Poitou 1981. Progress in Math., Vol. 22, Birkhäuser, Boston, MA, 1982, 281–307. MR **84g**:10041
19. R. C. Vaughan, *On Waring's problem for cubes*, J. Reine Angew. Math. **365** (1986), 122–170. MR **87j**:11103
20. ———, *On Waring's problem for smaller exponents. II*, Mathematika **33** (1986), 6–22. MR **87j**:11104
21. ———, *The Hardy-Littlewood method*, 2nd ed., Cambridge Tracts in Mathematics **125**, Cambridge University Press, Cambridge, U.K., 1997. MR **98a**:11133
22. Y. Wang, *Diophantine equations and inequalities in algebraic number fields*, Springer-Verlag, Berlin/Heidelberg/New York, 1991. MR **92a**:11036
23. T. D. Wooley, *New estimates for smooth Weyl sums*, J. London Math. Soc. (2) **51** (1995), 1–13. MR **96e**:11109

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, 395 UCB, BOULDER, COLORADO 80309

Current address: School of Mathematics, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540

E-mail address: freem@ias.edu

ON THE CANONICAL RINGS OF COVERS OF SURFACES OF MINIMAL DEGREE

FRANCISCO JAVIER GALLEGO AND BANGERE P. PURNAPRAJNA

ABSTRACT. In one of the main results of this paper, we find the degrees of the generators of the canonical ring of a regular algebraic surface X of general type defined over a field of characteristic 0, under the hypothesis that the canonical divisor of X determines a morphism φ from X to a surface of minimal degree Y . As a corollary of our results and results of Ciliberto and Green, we obtain a necessary and sufficient condition for the canonical ring of X to be generated in degree less than or equal to 2. We construct new examples of surfaces satisfying the hypothesis of our theorem and prove results which show that many a priori plausible examples cannot exist. Our methods are to exploit the \mathcal{O}_Y -algebra structure on $\varphi_*\mathcal{O}_X$. These methods have other applications, including those on Calabi-Yau threefolds. We prove new results on homogeneous rings associated to a polarized Calabi-Yau threefold and also prove some existence theorems for Calabi-Yau covers of threefolds of minimal degree. These have consequences towards constructing new examples of Calabi-Yau threefolds.

INTRODUCTION

The canonical models of surfaces of general type have attracted the attention of many geometers. The questions on projective normality and ring generators of the canonical ring are of particular interest. Kodaira [Kod] first proved that $|K_X^{\otimes m}|$ embeds a minimal surface of general type X as a projectively normal variety for all $m \geq 8$. This was later improved by Bombieri [Bo], who proved the same result if $m \geq 6$, and by Ciliberto [Ci], who lowered the bound to $m \geq 5$. We proved in [GP1] more general results on projective normality and higher syzygies for adjunction bundles for an algebraic surface. As a corollary of these results we recovered and improved the results of Bombieri and Ciliberto on projective normality, and extended them to higher syzygies. We also recovered and extended the results of Reid [R] on the ring generators of the canonical ring of a surface of general type.

An important class of minimal surfaces of general type comprises those whose canonical divisor is base-point-free. Surfaces with base-point-free canonical divisor fall naturally into two categories corresponding to the division of curves of genus greater than one into hyperelliptic and non-hyperelliptic: those whose canonical

Received by the editors July 5, 2002.

2000 *Mathematics Subject Classification.* Primary 14J29.

The first author was partially supported by MCT project number BFM2000-0621 and by UCM project number PR52/00-8862. The second author was partially supported by the General Research Fund of the University of Kansas at Lawrence. The first author is grateful for the hospitality of the Department of Mathematics of the University of Kansas at Lawrence.

morphism maps onto a surface of minimal degree and those whose canonical morphism does not map to such a surface. By a surface of minimal degree we mean a nondegenerate surface in projective space whose degree is equal to its codimension plus 1. The surfaces of minimal degree are classically known: they are (linear) \mathbf{P}^2 , the Veronese surface in \mathbf{P}^5 , and smooth rational scrolls or cones over a rational normal curve (see [EH]). Note that, even though we are drawing an analogy between hyperelliptic curves and surfaces of general type whose canonical morphism maps onto a surface of minimal degree, the theory is much harder for surfaces, because higher degree covers are involved. Examples of such covers are shown in Section 3.

In this paper we deal with surfaces of general type whose canonical morphism maps onto a surface of minimal degree. The surfaces whose canonical morphism does not map onto a surface of minimal degree have been studied by Ciliberto and Green (see [Ci] and [G]). Green and Ciliberto proved the following beautiful result regarding the generators of the canonical ring:

Let X be a regular surface of general type with base-point-free canonical divisor. Assume that the canonical morphism φ satisfies the following conditions:

- (1) φ does not map X generically $2 : 1$ onto the projective plane;
- (2) $\varphi(X)$ is not a surface of minimal degree (other than linear \mathbf{P}^2).

Then the canonical ring of X is generated in degree less than or equal to 2.

The surfaces of general type X whose canonical morphism φ maps X onto a surface of minimal degree Y have been studied by Horikawa (see [H1], [H2], [H3] and [H4]), Catanese [Ca], Konno [Kon], Mendes Lopes and Pardini [MP], among others, where they play a central role in the classification of surfaces of general type with small c_1^2 and in questions about degenerations and the moduli of surfaces of general type. The study of these surfaces has a direct bearing on the study of linear series on threefolds such as Calabi-Yau threefolds, as the results in [OP], and the authors' results in [GP2] and Section 4 of this article show.

The study of the canonical rings of these surfaces is carried out in Section 2. We determine the precise degrees of the generators of its canonical ring (see Theorem 2.1). The answer depends on the degree of φ and the degree of Y . As a corollary of our result and the result of Ciliberto and Green (see Corollary 2.8), we find that conditions (1) and (2) above characterize the regular surfaces of general type with base-point-free canonical bundle whose canonical ring is generated in degree less than or equal to 2. This result is surprising, because it contrasts with the situation for higher-dimensional varieties, which, as we show in [GP3], differs from the situation for surfaces. Indeed, we show in [GP3] that there is no higher-dimensional analogue of Corollary 2.8, and therefore there is no converse of Green's result (cf. [G], Theorem 3.9.3) for higher dimensional varieties of general type.

In Section 2 we explain how to use the \mathcal{O}_Y -algebra structure on $\varphi_*\mathcal{O}_X$ to find the multiplicative structure of the canonical ring of X . Even though we reduce the problem from a complicated variety to a simpler variety, as a surface Y of minimal degree is, there are certain difficulties that arise in the process. The proof of Theorem 2.1 involves the study of multiplication of global sections of line bundles on a surface X of general type. To do so we reduce the problem to Y , and this amounts to studying maps of multiplication of global sections of vector bundles instead of line bundles. This is the first difficulty. Moreover, the relation between the multiplication maps of global sections of line bundles on X and the multiplication maps of global sections of vector bundles on Y is governed by the

\mathcal{O}_Y -algebra structure on $\varphi_*\mathcal{O}_X$. However, we prove our results for a large class of surfaces X . So their canonical morphism φ might, and actually does, correspond to many, quite diverse \mathcal{O}_Y -algebra structures. This is in sharp contrast to hyperelliptic curves, where canonical morphisms correspond to degree 2 algebra structures, which are all quite similar and very easy to describe. The algebra structures arising from canonical morphisms of surfaces are much more complicated and harder to determine. This is the second difficulty one encounters, a difficulty that we are able to overcome in the context of this paper.

In Section 3 we construct new examples of surfaces of general type mapping to a surface of minimal degree and also recall some known ones. It is interesting to know in general what positive integers occur as degrees of the canonical morphism if the image is a surface of minimal degree. An answer to this question is helpful in finding new examples. Having this philosophy in mind, we prove results which show that some natural ways to construct examples do not work. For instance, it follows from the results in Section 3 that odd degree covers of smooth scrolls or cyclic covers of degree bigger than 3 of surfaces of minimal degree, induced by the canonical morphism, do not exist.

The results on surfaces of general type mentioned above have ramifications for Calabi-Yau threefolds. In Section 4 of this article we apply these results to obtain new results for a polarized Calabi-Yau threefold (X, B) with B a base-point-free and ample divisor. Among other things, we find out the degrees of the generators of the homogeneous ring associated to B , and we give a characterization of polarized Calabi-Yau threefolds (X, B) whose associated homogeneous ring is generated in degree less than or equal to 2. The construction of examples of Calabi-Yau threefolds has evoked interest in recent years. One of the important sources of these examples is to take covers of threefolds of minimal degree. In Section 4, we prove some existence theorems for Calabi-Yau covers of threefolds of minimal degree. For instance, from these results, which are more general, it follows that a Calabi-Yau cover of prime degree greater than 3 induced by a complete linear series cannot come from a group action.

We will expand on these ideas in two forthcoming articles, [GP3] and [GP5]. In the first we study the canonical ring of higher-dimensional varieties of general type whose canonical morphism maps onto a variety of minimal degree. One of the results in [GP3] shows that the converse of the theorem of Ciliberto and Green for surfaces proved in this article is false for higher-dimensional varieties of general type. In the second we carry out a detailed study of homogeneous rings associated to line bundles on trigonal curves.

1. PRELIMINARIES

Convention. *Throughout this article we will work over an algebraically closed field of characteristic 0.*

In this section we will recall some known facts about the push-forward of the structure sheaf of a variety by a flat, finite morphism. We summarize these facts below, and refer for the proof to [HM], Section 2.

Let X and Y be algebraic varieties. Let $\pi : X \rightarrow Y$ be a finite, flat morphism of degree n . We have the following facts:

1.1. The sheaf $\pi_*\mathcal{O}_X$ is a rank n , locally free sheaf on Y of algebras over \mathcal{O}_Y .

1.2. There exists a map

$$\frac{1}{n}\text{tr} : \pi_*\mathcal{O}_X \longrightarrow \mathcal{O}_Y$$

of sheaves of \mathcal{O}_Y -modules defined locally as follows: Given $\alpha \in \pi_*\mathcal{O}_X$, we consider the homomorphism of \mathcal{O}_Y -modules

$$\pi_*\mathcal{O}_X \xrightarrow{\alpha} \pi_*\mathcal{O}_X$$

induced by multiplication by α . Then we define $\frac{1}{n}\text{tr}(\alpha)$ as the trace of such a homomorphism divided by n .

1.3. $\frac{1}{n}\text{tr}$ is surjective; in fact, the map $\mathcal{O}_Y \hookrightarrow \pi_*\mathcal{O}_X$ induced by π is a section of $\frac{1}{n}\text{tr}$. Therefore the sequence

$$0 \longrightarrow E \longrightarrow \pi_*\mathcal{O}_X \xrightarrow{\frac{1}{n}\text{tr}} \mathcal{O}_Y \longrightarrow 0$$

splits. E is the kernel of $\frac{1}{n}\text{tr}$ and locally consists of the trace 0 elements of $\pi_*\mathcal{O}_X$. We will call E the *trace-zero module* of π .

1.4. $\pi_*(\mathcal{O}_X)$ is a sheaf of \mathcal{O}_Y -algebras; therefore, it has a multiplicative structure. Its multiplication map is an \mathcal{O}_Y -linear map

$$[\mathcal{O}_Y \oplus E] \otimes [\mathcal{O}_Y \oplus E] \longrightarrow \mathcal{O}_Y \oplus E$$

made of four components. The first component

$$\mathcal{O}_Y \otimes \mathcal{O}_Y \longrightarrow \mathcal{O}_Y \oplus E$$

is given by the multiplication in \mathcal{O}_Y , and therefore goes to \mathcal{O}_Y . The components

$$\mathcal{O}_Y \otimes E \longrightarrow \mathcal{O}_Y \oplus E,$$

$$E \otimes \mathcal{O}_Y \longrightarrow \mathcal{O}_Y \oplus E$$

are given by the left and right module structure of E over \mathcal{O}_Y , and therefore go to E . Finally, there is a fourth component

$$E \otimes E \longrightarrow \mathcal{O}_Y \oplus E$$

which factors through

$$S^2 E \longrightarrow \mathcal{O}_Y \oplus E,$$

for multiplication in $\pi_*\mathcal{O}_Y$ is commutative.

2. COVERS OF SURFACES OF MINIMAL DEGREE

Our purpose in this section is to study the generators of the canonical ring of certain surfaces of general type. Specifically, we are interested in studying those regular surfaces of general type whose canonical divisor is base-point-free and such that the image of the canonical morphism is a variety of minimal degree. We obtain the following result.

Theorem 2.1. *Let S be a regular surface of general type with at worst canonical singularities and such that its canonical bundle K_S is base-point-free. Let φ be the canonical morphism of S . Let n be the degree of φ and assume that the image of φ is a surface of minimal degree r . Then:*

1) *if $n = 2$ and $r = 1$ (i.e., if φ is generically $2 : 1$ onto \mathbf{P}^2), then the canonical ring of S is generated by its part of degree 1 and one generator in degree 4;*

2) *if $n \neq 2$ or $r \neq 1$, then the canonical ring of S is generated by its part of degree 1, $r(n - 2)$ generators in degree 2 and $r - 1$ generators in degree 3.*

The knowledge of how many linearly independent generators are needed in each degree is obtained from the knowledge of the image of the multiplication maps of global sections of powers of the canonical bundle. We study those multiplication maps by studying similar maps of a curve C in $|K_S|$. Thus we will first prove the following proposition.

Proposition 2.2. *Let C be a smooth, irreducible curve. Let θ be a base-point-free line bundle on C such that $\theta^{\otimes 2} = K_C$. Let π be the morphism induced by $|\theta|$, let n be the degree of π , and assume that $\pi(C)$ is a rational normal curve of degree r . Let $\beta(s, t)$ be the multiplication map*

$$H^0(\theta^{\otimes s}) \otimes H^0(\theta^{\otimes t}) \longrightarrow H^0(\theta^{\otimes s+t}), \text{ for all } s, t > 0.$$

The codimension of the image of $\beta(s, t)$ in $H^0(\theta^{\otimes s+t})$ is as follows:

- a) *If $r = 1$, the codimension is:*
 - a.1) $n - 2$, for $s = t = 1$,
 - a.2) 0 , for $s = 2, t = 1$, i.e., $\beta(2, 1)$ surjects,
 - a.3) 1 , for $s = 3, t = 1$,
 - a.4) 1 , for $s = t = 2$, $n = 2$ and 0 if $n > 2$,
 - a.5) 0 , for $s \geq 4, t = 1$, i.e., $\beta(s, 1)$ surjects for all $s \geq 4$.
- b) *If $r > 1$, the codimension is:*
 - b.1) $r(n - 2)$, for $s = t = 1$,
 - b.2) $r - 1$, for $s = 2, t = 1$,
 - b.3) 0 , for $s \geq 3, t = 1$, i.e., $\beta(s, 1)$ surjects for all $s \geq 3$.

Moreover, if $r = 1$ and $n = 2$, then the image of $\beta(2, 2)$ and the image of $\beta(3, 1)$ are equal.

In order to prove Proposition 2.2, we will use the following.

Lemma 2.3. *Let C , θ and π be as in the statement of Proposition 2.2. Then*

$$\pi_* \mathcal{O}_C = \mathcal{O}_{\mathbf{P}^1} \oplus (n - 2) \mathcal{O}_{\mathbf{P}^1}(-r - 1) \oplus \mathcal{O}_{\mathbf{P}^1}(-2r - 2).$$

Proof. Since the image of π is smooth and of dimension 1, π is flat. Then $\pi_* \mathcal{O}_C = \mathcal{O}_{\mathbf{P}^1} \oplus E$ as $\mathcal{O}_{\mathbf{P}^1}$ -modules, with E a vector bundle over \mathbf{P}^1 of rank $n - 1$. We now show that

$$E = (n - 2) \mathcal{O}_{\mathbf{P}^1}(-r - 1) \oplus \mathcal{O}_{\mathbf{P}^1}(-2r - 2).$$

We have $\pi_* \theta = \pi_* \mathcal{O}_C \otimes \mathcal{O}_{\mathbf{P}^1}(r)$ and $\pi_* K_C = \pi_* \mathcal{O}_C \otimes \mathcal{O}_{\mathbf{P}^1}(2r)$, by the projection formula. Any vector bundle over \mathbf{P}^1 splits; hence

$$\pi_* \mathcal{O}_C = \mathcal{O}_{\mathbf{P}^1} \oplus E = \mathcal{O}_{\mathbf{P}^1} \oplus \mathcal{O}_{\mathbf{P}^1}(a_1) \oplus \cdots \oplus \mathcal{O}_{\mathbf{P}^1}(a_{n-1}),$$

for some negative integers a_1, \dots, a_{n-1} (C is connected). Then $h^1(K_C) = 1$ implies that exactly one of the a_i 's, let us say a_{n-1} , satisfies $a_{n-1} + 2r = -2$. On the other hand, since π is induced by the complete linear series $|\theta|$, $h^0(\theta) = r + 1 = h^0(\mathcal{O}_{\mathbf{P}^1}(r))$; so $a_i + r \leq -1$ for all $1 \leq i \leq n - 2$. Finally, since the degree of θ is $g(C) - 1$, we have $h^1(\theta) = h^0(\theta) = r + 1$. Since $h^1(\mathcal{O}_{\mathbf{P}^1}(-r - 2)) = r + 1$, $a_i + r \geq -1$ for all $1 \leq i \leq n - 2$, and so $a_i + r = -1$ for all $1 \leq i \leq n - 2$. \square

(2.4) *Proof of Proposition 2.2.* In Lemma 2.3, we have completely determined the structure of $\pi_* \mathcal{O}_C$ as an $\mathcal{O}_{\mathbf{P}^1}$ -module. Now we look at the structure of $\pi_* \mathcal{O}_C$

as an $\mathcal{O}_{\mathbf{P}^1}$ -algebra. If $n = 2$, it is completely determined by the branch divisor of π on \mathbf{P}^1 , since in this case π is cyclic. If $n > 2$, we observe the following:

For some $1 \leq i, j \leq n - 2$, the projection of the map

$$(2.4.1) \quad \mathcal{O}_{\mathbf{P}^1}(a_i) \otimes \mathcal{O}_{\mathbf{P}^1}(a_j) \longrightarrow \pi_* \mathcal{O}_C \text{ to } \mathcal{O}_{\mathbf{P}^1}(-2r - 2)$$

is surjective; in fact, it is an isomorphism.

This is so because otherwise $\mathcal{O}_{\mathbf{P}^1} \oplus \mathcal{O}_{\mathbf{P}^1}(a_1) \oplus \cdots \oplus \mathcal{O}_{\mathbf{P}^1}(a_{n-2})$ would be an integral subalgebra of $\pi_* \mathcal{O}_C$, free over $\mathcal{O}_{\mathbf{P}^1}$ of rank $n - 1$. Then $n - 1$ should divide n , which is not possible if $n > 2$.

Now we will use our knowledge of $\pi_* \mathcal{O}_X$ to study the maps $\beta(s, r)$ which appear in the statement of the proposition. We will write β_s in place of $\beta(s, 1)$. Let $R_l = H^0(\theta^{\otimes l})$. Then, since $\theta = \pi^* \mathcal{O}_{\mathbf{P}^1}(r)$, by the projection formula,

$$R_1 = H^0(\mathcal{O}_{\mathbf{P}^1}(r)),$$

$$R_l = H^0(\mathcal{O}_{\mathbf{P}^1}(lr)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 1)) \oplus H^0(\mathcal{O}_{\mathbf{P}^1}((l - 2)r - 2)),$$

$$R_{l+1} = H^0(\mathcal{O}_{\mathbf{P}^1}((l + 1)r)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}(lr - 1)) \oplus H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 2)).$$

Therefore an element of R_l , i.e., a global section of $H^0(\theta^{\otimes l})$, is a sum of n components, one in each piece of the above decomposition of R_l . On the other hand, the product of an element of R_l belonging to one of the blocks with an element of R_1 is determined by the ring structure of $\mathcal{O}_{\mathbf{P}^1}$ and by the module structure of E . More precisely, the restriction of β_l to $H^0(\mathcal{O}_{\mathbf{P}^1}(lr)) \otimes H^0(\mathcal{O}_{\mathbf{P}^1}(r))$ maps, in fact isomorphically, onto $H^0(\mathcal{O}_{\mathbf{P}^1}((l + 1)r))$. The restriction of β_l to each of the blocks $H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 1)) \otimes H^0(\mathcal{O}_{\mathbf{P}^1}(r))$ maps to the corresponding $H^0(\mathcal{O}_{\mathbf{P}^1}(lr - 1))$. This restriction is 0 if $(l - 1)r - 1$ is negative and an isomorphism otherwise. Likewise, the restriction of β_l to $H^0(\mathcal{O}_{\mathbf{P}^1}((l - 2)r - 2)) \otimes H^0(\mathcal{O}_{\mathbf{P}^1}(r))$ goes to $H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 2))$, being 0 if $(l - 2)r - 2$ is negative and an isomorphism otherwise. Therefore it is crucial to tell which blocks of a given R_l are 0. We have

$$R_1 = H^0(\mathcal{O}_{\mathbf{P}^1}(r)),$$

$$R_2 = H^0(\mathcal{O}_{\mathbf{P}^1}(2r)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}(r - 1)),$$

and if $l \geq 3$,

$$R_l = H^0(\mathcal{O}_{\mathbf{P}^1}(lr)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 1)) \oplus H^0(\mathcal{O}_{\mathbf{P}^1}((l - 2)r - 2)).$$

All the direct summands appearing in the above formulae are nonzero, except $H^0(\mathcal{O}_{\mathbf{P}^1}((l - 2)r - 2))$ when $l = 3$ and $r = 1$ and $(n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}((l - 1)r - 1))$ for all l and all r when $n = 2$. We now determine the image of β_l . If $l = 1$, the image of β_1 is $H^0(\mathcal{O}_{\mathbf{P}^1}(2r))$, which has codimension $(n - 2)r$ in R_2 . If $l = 2$, the image of β_2 is $H^0(\mathcal{O}_{\mathbf{P}^1}(3r)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}(2r - 1))$, which has codimension $r - 1$ in R_3 . If $l = 3$ and $r \geq 2$ or if $l \geq 4$, the image of β_l is all R_l , i.e., β_l surjects. All this proves a.1), a.2), a.5) and b). If $r = 1$, the image of $\beta(3, 1)$ is $H^0(\mathcal{O}_{\mathbf{P}^1}(4r)) \oplus (n - 2)H^0(\mathcal{O}_{\mathbf{P}^1}(3r - 1))$, which has codimension 1 in R_4 . This proves a.3). If $r = 1$ and $n = 2$, the image of $\beta(2, 2)$ is $H^0(\mathcal{O}_{\mathbf{P}^1}(4r))$, which has codimension 1 in R_4 . This proves the first claim in a.4) and the last sentence of Proposition 2.2. Finally, if $n > 2$, recall (see 2.4.1) that for some $1 \leq i, j \leq n - 2$, the projection of the map

$$\mathcal{O}_{\mathbf{P}^1}(a_i) \otimes \mathcal{O}_{\mathbf{P}^1}(a_j) \longrightarrow \pi_* \mathcal{O}_C$$

to $\mathcal{O}_{\mathbf{P}^1}(-4)$ is surjective; in fact, it is an isomorphism. Then, if $n > 2$, the image of $\beta(2, 2)$ is all R_4 . This proves the second part of a.4). \square

Remark 2.5. Note that $\theta^{\otimes 2} = K_C$. Then a proof of a.4), alternate to the one given above, can be obtained from Noether's theorem and from the base-point-free pencil trick. The way in which Noether's theorem is related to the algebra structure of $\pi_*\mathcal{O}_C$ is shown in [GP4], where we will give a different, simple proof of this classical result for a general curve in M_g^1 .

From Proposition 2.2 we obtain the following.

Corollary 2.6. *Let C be a smooth curve. Let θ be a base-point-free line bundle on C such that $\theta^{\otimes 2} = K_C$. Let π be the morphism induced by $|\theta|$, let n be the degree of π and assume that $\pi(C)$ is a rational normal curve of degree r . Let R be $\bigoplus_{l=0}^{\infty} H^0(\theta^{\otimes l})$. Then:*

1) *if $r = 1$ and $n = 2$, the ring R is generated by its part of degree 1 and one generator in degree 4;*

2) *if $r = 1$ and $n > 2$, the ring R is generated by its part of degree 1 and $n - 2$ generators in degree 2;*

3) *if $r > 1$, the ring R is generated by its part of degree 1, $r(n - 2)$ generators in degree 2 and $r - 1$ generators in degree 3.*

Proof. To know in what degrees we need generators, we look at the maps $\beta(s, t)$ of multiplication of sections. Precisely the number of generators needed in degree $l + 1$ is the codimension in R_{l+1} of the sum of the images of $\beta(l, 1)$, $\beta(l - 1, 2), \dots, \beta(\lfloor \frac{l+1}{2} \rfloor, \lceil \frac{l+1}{2} \rceil)$. In particular, R is generated in degree less than or equal to l if β_k surjects for all $k \geq l$. Thus 1) follows from part a) of Proposition 2.2 and from the fact that the images of $\beta(3, 1)$ and $\beta(2, 2)$ are equal. 2) follows likewise from part a) of Proposition 2.2 (note that in this case $\beta(2, 2)$ surjects). Finally, 3) follows from part b) of Proposition. \square

(2.7) *Proof of Theorem 2.1.* The proof rests on Proposition 2.2. The idea is "to lift" the generators of R to the canonical ring of S . Let us define

$$H^0(K_S^{\otimes s}) \otimes H^0(K_S^{\otimes r}) \xrightarrow{\alpha(s,t)} H^0(K_S^{\otimes s+t}),$$

and let us denote $\alpha(s, 1)$ as α_s . As in the case of R , the images of $\alpha(s, t)$ will tell us the generators of each graded piece of the canonical ring of S . In fact, it will suffice to prove the following:

(a) If $r = 1$ and $n = 2$, α_l surjects for all $l \geq 1$, except if $l = 3$. The images of $\alpha_3 = \alpha(3, 1)$ and $\alpha(2, 2)$ are equal and have codimension 1 in $H^0(K_S^{\otimes 4})$.

(b) If $r = 1$ and $n > 2$, α_l surjects for all $l \geq 1$, except if $l = 1, 3$. The image of α_1 has codimension $n - 2$ in $H^0(K_S^{\otimes 2})$. The map $\alpha(2, 2)$ is surjective.

(c) If $r \geq 2$, α_l is surjective if $l \geq 3$. The image of α_1 has codimension $r(n - 2)$ in $H^0(K_S^{\otimes 2})$. The image of α_2 has codimension $r - 1$ in $H^0(K_S^{\otimes 3})$.

Thus we proceed to prove (a), (b), (c). Recall that $Y = \varphi(S)$ is an irreducible variety of minimal degree and, in particular, normal. On the other hand, the locus of the points of Y with non-finite fibers has codimension 2. Thus, using Bertini's theorem, we can choose a smooth curve C of $|K_S|$ such that the restriction of the canonical morphism of S to C is finite (and flat) onto a smooth rational normal curve of degree r . Let us denote by θ the restriction of K_S to C . By adjunction,

$K_C = \theta^{\otimes 2}$. Since K_S is base-point-free, so is θ . Finally, since $H^1(\mathcal{O}_X) = 0$, π is induced by the complete linear series $|\theta|$, and therefore C , θ and π satisfy the hypothesis of Proposition 2.2.

We prove first the statements in (a), (b) and (c) regarding the maps α_l . Consider the following commutative diagram:

$$\begin{array}{ccccc} H^0(K_S^{\otimes l}) \otimes H^0(\mathcal{O}_S) & \hookrightarrow & H^0(K_S^{\otimes l}) \otimes H^0(K_S) & \twoheadrightarrow & H^0(K_S^{\otimes l}) \otimes H^0(\theta) \\ \downarrow & & \downarrow \alpha_l & & \downarrow \\ H^0(K_S^{\otimes l}) & \hookrightarrow & H^0(K_S^{\otimes l+1}) & \twoheadrightarrow & H^0(\theta^{\otimes l+1}) \end{array}$$

The rightmost horizontal arrows are surjective because $H^1(\mathcal{O}_S) = 0$, by Serre duality and by Kawamata-Viehweg vanishing. The left vertical arrow trivially surjects. The right vertical arrow is the composition of the map $H^0(K_S^{\otimes l}) \otimes H^0(\theta) \rightarrow H^0(\theta^{\otimes l}) \otimes H^0(\theta)$, which is surjective for all $l \geq 1$ again because $H^1(\mathcal{O}_S) = 0$, by Serre duality and by Kawamata-Viehweg vanishing, and the map β_l of multiplication of global sections on C , studied in Proposition 2.2. Then it follows from chasing the diagram that the map $H^0(K_S^{\otimes l+1}) \rightarrow H^0(\theta^{\otimes l+1})$ maps the image of α_l onto the image of β_l , and that the codimension of the image of β_l in $H^0(\theta^{\otimes l+1})$ is equal to the codimension of the image of α_l in $H^0(K_S^{\otimes l+1})$. This, together with Proposition 2.2, a.1, a.2, a.3, a.5 and b, proves the claims in (a), (b) and (c) concerning the codimensions of the images of the maps α_l .

Thus the only things left to prove are the claims about $\alpha(2, 2)$ when $r = 1$. We consider the commutative diagram

$$\begin{array}{ccccc} H^0(K_S^{\otimes 2}) \otimes H^0(K_S) & \hookrightarrow & H^0(K_S^{\otimes 2}) \otimes H^0(K_S^{\otimes 2}) & \twoheadrightarrow & H^0(K_S^{\otimes 2}) \otimes H^0(\theta^{\otimes 2}) \\ \downarrow \alpha_2 & & \downarrow \alpha(2,2) & & \downarrow \\ H^0(K_S^{\otimes 3}) & \hookrightarrow & H^0(K_S^{\otimes 4}) & \twoheadrightarrow & H^0(\theta^{\otimes 4}) \end{array}$$

The rightmost horizontal arrows are surjective because $H^1(\mathcal{O}_S) = 0$ and by Serre duality, and by Kawamata-Viehweg vanishing. The left vertical arrow surjects, as we have already proven. The right vertical arrow is the composition of the map $H^0(K_S^{\otimes 2}) \otimes H^0(\theta^{\otimes 2}) \rightarrow H^0(\theta^{\otimes 2}) \otimes H^0(\theta^{\otimes 2})$, which is surjective because S is regular and by Serre duality, and the map $\beta(2, 2)$ of multiplication of global sections on C . Then it follows from chasing the diagram that the map $H^0(K_S^{\otimes 4}) \rightarrow H^0(\theta^{\otimes 4})$ maps the image of $\alpha(2, 2)$ onto the image of $\beta(2, 2)$, and that the codimension of the image of $\beta(2, 2)$ in $H^0(\theta^{\otimes 4})$ is equal to the codimension of the image of $\alpha(2, 2)$ in $H^0(K_S^{\otimes 4})$. On the other hand, we know that the image of $\beta(2, 2)$ and the image of $\beta_3 = \beta(3, 1)$ are equal of codimension 1 in $H^0(\theta^{\otimes 4})$, if $r = 1$ and $n = 2$. Thus we conclude that the images of $\alpha(3, 1)$ and $\alpha(2, 2)$ in $H^0(K_S^{\otimes 4})$ are also equal and of codimension 1. Finally, if $r = 1$ and $n > 2$, $\beta(2, 2)$ surjects by Proposition 2.2.a.4. Thus we conclude that if $r = 1$ and $n > 2$, then $\alpha(2, 2)$ surjects. \square

Theorem 2.1 complements known results on generation of the canonical ring of smooth, regular surfaces of general type. Ciliberto and Green (cf. [G], Theorem 3.9.3, and [Ci]) proved that, given a smooth surface of general type with $h^1(\mathcal{O}_S) = 0$ and K_S globally generated and φ being the canonical morphism, a sufficient condition for the canonical ring of S to be generated in degree less than or equal to 2 is that:

- (1) φ does not map S generically $2 : 1$ onto \mathbf{P}^2 , and
- (2) $\varphi(S)$ is not a surface of minimal degree other than linear \mathbf{P}^2 .

As a corollary of the Ciliberto and Green result and of Theorem 2.1, we obtain the following:

Corollary 2.8. *Let S be a smooth regular surface of general type and such that K_S is globally generated. Let φ be the canonical morphism of S . The canonical ring of S is generated in degree less than or equal to 2 if and only if*

- (1) φ does not map S generically $2 : 1$ onto \mathbf{P}^2 , and
- (2) $\varphi(S)$ is not a surface of minimal degree other than linear \mathbf{P}^2 .

3. EXAMPLES OF SURFACES OF GENERAL TYPE

In this section we construct some new examples of surfaces of general type that satisfy the hypothesis of Theorem 2.1. The easiest way one could think of producing examples would be to build suitable cyclic covers of surfaces of minimal degree. However, as the next proposition shows, only low degree cyclic covers can be induced by the canonical morphism of a regular surface. So we have to employ other means to construct these examples.

Proposition 3.1. *Let X be a surface of general type with at worst canonical singularities and with base-point-free canonical bundle. Assume that the complete canonical series of X restricts to a complete linear series on a general member of the canonical series (e.g., if X is regular). Let $\varphi : X \rightarrow Y$ be the canonical morphism to a surface of minimal degree. Let n be the degree of φ . Let U be a smooth open set of Y whose complement has codimension 2 and let L be a line bundle on U . Assume that*

$$(\varphi_* \mathcal{O}_X)|_U = \mathcal{O}_U \oplus L^{-1} \oplus \cdots \oplus L^{\otimes 1-n}.$$

Then $n = 2$ or 3 .

Proof. Let H be a general hyperplane section of Y contained in U and let C be the inverse image of H by φ . Then C is a smooth irreducible member of $|K_X|$ and H is a smooth rational normal curve. By assumption the morphism $\varphi|_C : C \rightarrow H$ is induced by the complete linear series of a line bundle θ . By adjunction $\theta^{\otimes 2} = K_C$. Thus C , θ and $\varphi|_C$ satisfy the hypothesis of Lemma 2.3, and

$$(\varphi|_C)_* \mathcal{O}_C = \mathcal{O}_{\mathbf{P}^1} \oplus (n-2)\mathcal{O}_{\mathbf{P}^1}(-r-1) \oplus \mathcal{O}_{\mathbf{P}^1}(-2r-2).$$

On the other hand, $(\varphi|_C)_* \mathcal{O}_C$ is equal to the restriction of $\varphi_* \mathcal{O}_X$ to H , and hence

$$(\varphi|_C)_* \mathcal{O}_C = \mathcal{O}_{\mathbf{P}^1} \oplus L'^{-1} \oplus \cdots \oplus L'^{1-n},$$

where L' is the restriction of L to H . The only way in which $(\varphi|_C)_* \mathcal{O}_C$ can have these two splittings is when $n = 2$ or 3 . \square

Corollary 3.2. *Let X be a regular surface of general type with at worst canonical singularities and with base-point-free canonical bundle. Let Y be the image of X by its canonical morphism $X \xrightarrow{\varphi} Y$. If Y is a surface of minimal degree and φ is a cyclic cover, then the degree of φ is 2 or 3.*

The next proposition also rules out many possible examples of covers of odd degree:

Proposition 3.3. *Let X be a surface of general type with at worst canonical singularities whose canonical divisor is base-point-free. Let φ be a morphism induced by a subseries of $|K_X|$. If φ is generically finite onto a smooth scroll $Y \subset \mathbf{P}^N$, then the degree of φ is even. In particular, there are not generically finite covers of odd degree of smooth rational normal scrolls induced by subseries of K_X .*

Proof. Let f be a fiber of Y and let C be a section of Y . Let $-d = C^2$. Since Y is a scroll, its hyperplane section is linearly equivalent to $C + mf$, for some integer m . Then $K_X = \varphi^*(C + mf)$. Then $\deg \varphi = (\varphi^*f) \cdot (\varphi^*C) = (\varphi^*f) \cdot (K_X - m\varphi^*f) = (\varphi^*f) \cdot (K_X + \varphi^*f)$, which is an even number. \square

Now we construct some examples of regular minimal surfaces X whose canonical morphism φ maps onto a variety of minimal degree, and also mention known ones relevant to this paper.

The cases when φ is a generically finite morphism and has degree 2 or 3 have been completely studied by Horikawa and Konno (see [H1], Theorem 1.6, [H2], Theorem 2.3.I, [H3], Theorem 4.1 and [Kon], Lemma 2.2 and Theorem 2.3; see also Mendes Lopes and Pardini, [MP]). As it turns out, there exist generically double covers of linear \mathbf{P}^2 , the Veronese surface, smooth rational normal scrolls $S(a, b)$ with $b \leq 4$, and cones over rational normal curves of degree 2, 3 and 4 and generically triple covers of \mathbf{P}^2 (in particular, cyclic triple covers of \mathbf{P}^2 ramified along a sextic with suitable singularities) and of the cones over rational normal curves of degree 2 and 3. Horikawa (see [H4], Theorem 2.1) also describes all generically finite quadruple covers $X \xrightarrow{\varphi} Y$, where X is a smooth, minimal regular surface, φ is the canonical morphism of X , and Y is linear \mathbf{P}^2 .

The examples of Horikawa and Konno just reviewed are examples of covers of degree less than or equal to 3 of surfaces of minimal degree and quadruple covers of \mathbf{P}^2 . We now construct three new sets of examples of regular surfaces of general type that are quadruple covers of surfaces of minimal degree under the canonical morphism. These examples are 4:1 covers of smooth rational normal scrolls isomorphic to the Hirzebruch surfaces \mathbf{F}_0 and \mathbf{F}_1 , and of quadric cones in \mathbf{P}^3 .

Example 3.4. *We construct finite quadruple covers $X \xrightarrow{\varphi} Y$, where X is a smooth minimal regular surface of general type, φ is the canonical morphism of X , and Y is a smooth rational scroll $S(m, m)$, $m \geq 1$.*

Let f be a fiber of one of the fibrations of \mathbf{P}^1 and let f' be a fiber of the other fibration. Then Y is $\mathbf{P}^1 \times \mathbf{P}^1$, and it is embedded in \mathbf{P}^{2m+1} by $|f + mf'|$ or by $|f' + mf|$. If Y is embedded by $|f + mf'|$, let a_1, a_2, b_1 and b_2 satisfy the following:

$$\begin{aligned} &\text{either } a_1 = 1, a_2 = 2, b_1 = m + 1 \text{ and } b_2 = 1, \\ &\text{or } a_1 = 2, a_2 = 1, b_1 = 1 \text{ and } b_2 = m + 1. \end{aligned}$$

If Y is embedded by $|f' + mf|$, let a_1, a_2, b_1 and b_2 satisfy the following:

$$\begin{aligned} &\text{either } b_1 = 1, b_2 = 2, a_1 = m + 1 \text{ and } a_2 = 1, \\ &\text{or } b_1 = 2, b_2 = 1, a_1 = 1 \text{ and } a_2 = m + 1. \end{aligned}$$

For $i = 1, 2$, let D_i be a smooth divisor linearly equivalent to $2(a_i f + b_i f')$ such that D_1 and D_2 intersect at $D_1 \cdot D_2$ distinct points. Those divisors exist because by the choices of a_1, a_2, b_1 and b_2 , both $2(a_1 f + b_1 f')$ and $2(a_2 f + b_2 f')$ are very ample. Let $X' \xrightarrow{\varphi_1} Y$ be the double cover of Y ramified along D_1 . Since D_1 is

smooth, so is X' . Let D'_2 be the inverse image in X' of D_2 by φ_1 . Since D_2 is smooth and meets D_1 at distinct points, D'_2 is also smooth. Let $X \xrightarrow{\varphi_2} X'$ be the double cover of X' ramified along D'_2 . Since X' and D'_2 are both smooth, so is X . Let $\varphi = \varphi_1 \circ \varphi_2$. Now we will show that X is a regular surface of general type, that $K_X = \varphi^* \mathcal{O}_Y(1)$, and that φ is induced by the complete canonical series of X . First we find out the structure of $\varphi_* \mathcal{O}_X$ as a module over \mathcal{O}_Y . Recall that $\varphi_{2*} \mathcal{O}_X = \mathcal{O}_{X'} \oplus \varphi_1^* \mathcal{O}_Y(-a_2 f - b_2 f')$. Then

$$\varphi_* \mathcal{O}_X = \varphi_{1*} \mathcal{O}_{X'} \oplus \varphi_{1*}(\varphi_1^* \mathcal{O}_Y(-a_2 f - b_2 f')).$$

Since $\varphi_{1*} \mathcal{O}_{X'} = \mathcal{O}_Y \oplus \mathcal{O}_Y(-a_1 f - b_1 f')$, then by the projection formula we have

$$\varphi_* \mathcal{O}_X = \mathcal{O}_Y \oplus \mathcal{O}_Y(-a_1 f - b_1 f') \oplus \mathcal{O}_Y(-a_2 f - b_2 f') \oplus \mathcal{O}_Y(-(a_1 + a_2)f - (b_1 + b_2)f').$$

We see now that X is regular. Recall that $H^1(\mathcal{O}_X) = H^1(\varphi_* \mathcal{O}_X)$. Our choice of a_1, a_2, b_1 and b_2 implies that $a_1 f + b_1 f'$ and $a_2 f + b_2 f'$ are both very ample; thus, by Kodaira vanishing,

$$\begin{aligned} H^1(\mathcal{O}_Y(-a_1 f - b_1 f')) &= H^1(\mathcal{O}_Y(-a_2 f - b_2 f')) \\ &= H^1(\mathcal{O}_Y(-(a_1 + a_2)f - (b_1 + b_2)f')) = 0. \end{aligned}$$

Then, since $H^1(\mathcal{O}_Y)$ also vanishes, so do $H^1(\varphi_* \mathcal{O}_X)$ and $H^1(\mathcal{O}_X)$. We now compute K_X . Since φ_2 is a double cover ramified along D'_2 ,

$$K_X = \varphi_2^*(K_{X'} \otimes \varphi_1^*(\mathcal{O}_Y(a_2 f + b_2 f'))).$$

For a similar reason,

$$K_{X'} = \varphi_1^*(K_Y \otimes \mathcal{O}_Y(a_1 f + b_1 f')).$$

Then

$$K_X = \varphi^*(K_Y \otimes \mathcal{O}_Y((a_1 + a_2)f + (b_1 + b_2)f')).$$

Since $K_Y = \mathcal{O}_Y(-2f - 2f')$, it follows again from the choices of a_1, a_2, b_1 and b_2 that $K_X = \varphi^* \mathcal{O}_Y(1)$. Finally, to see that φ is induced by the complete canonical linear series of X , we compute $H^0(K_X)$. We do the computation in the case $\mathcal{O}_Y(1) = \mathcal{O}_Y(f + mf')$. The case $\mathcal{O}_Y(1) = \mathcal{O}_Y(mf + f')$ is analogous. Since $K_X = \varphi^* \mathcal{O}_Y(1)$,

$$\begin{aligned} H^0(K_X) &= H^0(\mathcal{O}_Y(1)) \oplus H^0(\mathcal{O}_Y((1 - a_1)f + (m - b_1)f')) \\ &\quad \oplus H^0(\mathcal{O}_Y((1 - a_2)f + (m - b_2)f')) \\ &\quad \oplus H^0(\mathcal{O}_Y((1 - a_1 - a_2)f + (m - b_1 - b_2)f')). \end{aligned}$$

Again, by the choices of a_1, a_2, b_1 and b_2 , the last three direct sums of the above expression are 0. So φ is indeed induced by the complete canonical series of X .

Example 3.5. We construct finite quadruple covers $X \xrightarrow{\varphi} Y$, where X is a smooth regular surface of general type with base-point-free canonical bundle, φ is the canonical morphism of X , and Y is a smooth rational scroll $S(m-1, m)$, $m \geq 2$.

Let C_0 be the minimal section of \mathbf{F}_1 and let f be one of the fibers. Then Y is \mathbf{F}_1 , and it is embedded in \mathbf{P}^{2m} by $|C_0 + mf|$. Let a_1, a_2, b_1 and b_2 satisfy the following:

$$\begin{aligned} &\text{either } a_1 = 1, a_2 = 2, b_1 = m + 1 \text{ and } b_2 = 2, \\ &\text{or } a_1 = 2, a_2 = 1, b_1 = 2 \text{ and } b_2 = m + 1. \end{aligned}$$

For $i = 1, 2$, let D_i be a smooth divisor linearly equivalent to $2(a_i C_0 + b_i f)$, such that D_1 and D_2 intersect at $D_1 \cdot D_2$ distinct points. The fact that such divisors exist

follows from our choice of a_1, a_2, b_1 and b_2 , which implies that of the linear systems of D_1 and D_2 , one is very ample, and the other is base-point-free. Let $X' \xrightarrow{\varphi_1} Y$ be the double cover of Y ramified along D_1 . Since D_1 is smooth, so is X' . Let D'_2 be the inverse image in X' of D_2 by φ_1 . Since D_2 is smooth and meets D_1 transversally, D'_2 is also smooth. Let $X \xrightarrow{\varphi_2} X'$ be the double cover of X' ramified along D'_2 . Since X' and D'_2 are both smooth, so is X . Let $\varphi = \varphi_1 \circ \varphi_2$. Now we will show that X is a regular surface of general type, that $K_X = \varphi^* \mathcal{O}_Y(1)$, and that φ is induced by the complete canonical series of X . First we find the structure of $\varphi_* \mathcal{O}_X$ as a module over \mathcal{O}_Y . Recall that $\varphi_{2*} \mathcal{O}_X = \mathcal{O}_{X'} \oplus \varphi_1^* \mathcal{O}_Y(-a_2 C_0 - b_2 f)$. Then

$$\varphi_* \mathcal{O}_X = \varphi_{1*} \mathcal{O}_{X'} \oplus \varphi_{1*}(\varphi_1^* \mathcal{O}_Y(-a_2 C_0 - b_2 f)) .$$

Since $\varphi_{1*} \mathcal{O}_{X'} = \mathcal{O}_Y \oplus \mathcal{O}_Y(-a_1 C_0 - b_1 f)$, then by the projection formula we have

$$\begin{aligned} \varphi_* \mathcal{O}_X &= \mathcal{O}_Y \oplus \mathcal{O}_Y(-a_1 C_0 - b_1 f) \oplus \mathcal{O}_Y(-a_2 C_0 - b_2 f) \\ &\quad \oplus \mathcal{O}_Y(-(a_1 + a_2)C_0 - (b_1 + b_2)f) . \end{aligned}$$

We see now that X is regular. Recall that $H^1(\mathcal{O}_X) = H^1(\varphi_* \mathcal{O}_X)$. Our choices of a_1, a_2, b_1 and b_2 imply that $a_1 C_0 + b_1 f$ and $a_2 C_0 + b_2 f$ are both base-point-free and big divisors; thus, by Kawamata-Viehweg vanishing,

$$\begin{aligned} H^1(\mathcal{O}_Y(-a_1 C_0 - b_1 f)) &= H^1(\mathcal{O}_Y(-a_2 C_0 - b_2 f)) \\ &= H^1(\mathcal{O}_Y(-(a_1 + a_2)C_0 - (b_1 + b_2)f)) = 0 . \end{aligned}$$

Then, since $H^1(\mathcal{O}_Y)$ also vanishes, so does $H^1(\varphi_* \mathcal{O}_X)$ and therefore $H^1(\mathcal{O}_X)$. We now compute K_X . Since φ_2 is a double cover ramified along D'_2 ,

$$K_X = \varphi_2^*(K_{X'} \otimes \varphi_1^*(\mathcal{O}_Y(a_2 C_0 + b_2 f))) .$$

For a similar reason,

$$K_{X'} = \varphi_1^*(K_Y \otimes \mathcal{O}_Y(a_1 C_0 + b_1 f)) .$$

Then

$$K_X = \varphi^*(K_Y \otimes \mathcal{O}_Y((a_1 + a_2)C_0 + (b_1 + b_2)f)) .$$

Since $K_Y = \mathcal{O}_Y(-2C_0 - 3f)$, it follows from our choice of a_1, a_2, b_1 and b_2 that $K_X = \varphi^* \mathcal{O}_Y(1)$. Finally, to see that φ is induced by the complete canonical linear series of X , we compute $H^0(K_X)$. Since $K_X = \varphi^* \mathcal{O}_Y(1)$,

$$\begin{aligned} H^0(K_X) &= H^0(\mathcal{O}_Y(1)) \oplus H^0(\mathcal{O}_Y((1 - a_1)C_0 + (m - b_1)f)) \\ &\quad \oplus H^0(\mathcal{O}_Y((1 - a_2)C_0 + (m - b_2)f)) \\ &\quad \oplus H^0(\mathcal{O}_Y((1 - a_1 - a_2)C_0 + (m - b_1 - b_2)f)) . \end{aligned}$$

Again, by the choices of a_1, a_2, b_1 and b_2 , the last three direct sums of the above expression are 0. So φ is indeed induced by the complete canonical series of K_X .

Remark 3.6. With the same arguments, if one allows certain mild singularities in D_1 and D'_2 , then one can construct examples of covers of \mathbf{F}_0 and \mathbf{F}_1 with at worst canonical singularities.

Finally, we construct an example of a quadruple cover of a singular surface of minimal degree.

Example 3.7. We construct an example of a smooth, generically finite, quadruple cover $X \xrightarrow{\varphi} Z$ of the quadric cone Z in \mathbf{P}^3 , where X is a regular surface of general type whose canonical divisor is base-point-free, and φ is its canonical morphism.

Let $Y = \mathbf{F}_2$. Let C_0 be the minimal section of Y and let f be a fiber of Y . Let D_1 be a smooth divisor on Y , linearly equivalent to $2C_0 + 6f$ and meeting C_0 transversally. Let D_2 be a smooth divisor on Y , linearly equivalent to $3C_0 + 6f$ and meeting D_1 transversally. Such divisors D_1 and D_2 exist, because $2C_0 + 6f$ is very ample and $3C_0 + 6f$ is base-point-free. Note also that, since $(3C_0 + 6f) \cdot C_0 = 0$, C_0 and D_2 do not meet. Let $X' \xrightarrow{\varphi_1} Y$ be the double cover of Y along D_1 . Since D_1 is smooth, so is X' . Since D_1 meets C_0 at two distinct points, the pullback C'_0 of C_0 by φ_1 is a smooth line with self-intersection -4 . Let D'_2 be the pullback of D_2 by φ_1 . Since D_1 and D_2 meet transversally, D'_2 is smooth, and since D_2 and C_0 do not meet, neither do D'_2 and C'_0 . Let L'_2 be the pullback of $2C_0 + 3f$ by φ_1 . Let $X \xrightarrow{\varphi_2} X'$ be the double cover of X' along $D'_2 \cup C'_0$. Since $D'_2 \cup C'_0$ is smooth, so is X . Let $\varphi = \varphi_1 \circ \varphi_2$. Then

$$(3.7.1) \quad \begin{aligned} \varphi_* \mathcal{O}_X &= \varphi_{1*} \varphi_{2*} \mathcal{O}_X = \varphi_{1*} (\mathcal{O}_{X'} \oplus L'^{*}_2) \\ &= \mathcal{O}_Y \oplus \mathcal{O}_Y(-C_0 - 3f) \oplus \mathcal{O}_Y(-2C_0 - 3f) \oplus \mathcal{O}_Y(-3C_0 - 6f). \end{aligned}$$

Since $C_0 + 3f$ and $3C_0 + 6f$ are big and base-point-free, by Kawamata-Viehweg vanishing and Serre duality, $H^1(\mathcal{O}_Y(-C_0 - 3f)) = H^1(\mathcal{O}_Y(-3C_0 - 6f)) = 0$. By Serre duality, $H^1(\mathcal{O}_Y(-2C_0 - 3f)) = H^1(\mathcal{O}_Y(-f))^* = 0$. Then, since $H^1(\mathcal{O}_Y) = 0$, X is regular. Arguing as in Examples 3.4 and 3.5, we see that

$$(3.7.2) \quad K_X = \varphi^*(K_Y \otimes \mathcal{O}_Y(3C_0 + 6f)) = \varphi^* \mathcal{O}_Y(C_0 + 2f).$$

Now we compute $H^0(K_X)$. Using the projection formula and (3.7.1) and (3.7.2), we obtain that

$$\begin{aligned} H^0(K_X) &= H^0(\mathcal{O}_Y(C_0 + 2f)) \oplus H^0(\mathcal{O}_Y(-f)) \oplus H^0(\mathcal{O}_Y(-C_0 - f)) \\ &\quad \oplus H^0(\mathcal{O}_Y(-2C_0 - 4f)) = H^0(\mathcal{O}_Y(C_0 + 2f)). \end{aligned}$$

Thus the canonical morphism of X is the composition of φ and the morphism $Y \xrightarrow{\phi} Z \subset \mathbf{P}^3$, induced by the complete linear series of $C_0 + 2f$. Since ϕ contracts C_0 , the canonical morphism of X is not finite, but it is generically finite of degree 4 onto Z , which is a surface of minimal degree, as we wanted.

On the other hand, if C''_0 is the pullback of C_0 by φ , then C''_0 is a smooth line with self-intersection -2 . Thus the morphism $\phi \circ \varphi$ also factors as $\varphi' \circ \psi$, where

$$X \xrightarrow{\psi} \overline{X}$$

is the morphism from X to its canonical model \overline{X} and

$$\overline{X} \xrightarrow{\varphi'} Z$$

is the canonical morphism of \overline{X} . Thus φ' is an example of a finite, $4 : 1$ canonical morphism from a regular surface of general type with canonical singularities onto a singular surface of minimal degree.

4. APPLICATIONS TO CALABI-YAU THREEFOLDS

The results proved in Sections 2 and 3 have ramifications for Calabi-Yau threefolds. Recall that if X is a Calabi-Yau threefold and B is a big and base-point-free divisor, a general member of $|B|$ is a surface of general type. Then the geometry and properties of surfaces of general type are directly related to those of Calabi-Yau threefolds. Concretely, the results we have obtained in Section 2 on the canonical ring of surfaces of general type can be “lifted” to achieve analogous results for Calabi-Yau threefolds in a way similar to the way in which our study of rings of curves allowed us to obtain results for surfaces of general type. On the other hand, constructing examples of Calabi-Yau threefolds has attracted the attention of geometers in recent years. One of the important sources for these examples is precisely to take covers of varieties of minimal degree. Proposition 3.1, Corollary 3.2 and Proposition 3.3 tell us features of generically finite covers of surfaces of minimal degree induced by the canonical morphism. We will see how these features pass on to generically finite morphisms from Calabi-Yau threefolds to threefolds of minimal degree, and, as a consequence, we will obtain, among other things, that many a priori possible examples of Calabi-Yau threefolds cannot exist.

We start with the Calabi-Yau threefold analog of Theorem 2.1:

Theorem 4.1. *Let X be a Calabi-Yau threefold with at worst canonical singularities, and let B be a big and base-point-free line bundle on X . Let φ be the morphism induced by $|B|$. Let n be the degree of φ , and assume that the image of φ is a variety of minimal degree r . Then :*

- 1) *If $n = 2$ and $r = 1$ (i.e., if φ is generically $2 : 1$ onto \mathbf{P}^3), the canonical ring of X is generated by its part of degree 1 and one generator in degree 4.*
- 2) *If $n \neq 2$ or $r \neq 1$, the canonical ring of X is generated by its part of degree 1, $r(n - 2)$ generators in degree 2 and $r - 1$ generators in degree 3.*

Sketch of proof. The proof follows the same lines as the proof of Theorem 2.1. Let us define

$$H^0(B^{\otimes s}) \otimes H^0(B^{\otimes r}) \xrightarrow{\gamma(s,t)} H^0(B^{\otimes s+t}),$$

and denote $\gamma(s, 1)$ as γ_s . The images of $\gamma(s, t)$ will tell us the generators of each graded piece of the ring $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$. In fact, it would suffice to prove the following:

- (a) If $r = 1$ and $n = 2$, γ_l surjects for all $l \geq 1$, except if $l = 3$. The images of $\gamma_3 = \gamma(3, 1)$ and $\gamma(2, 2)$ are equal and have codimension 1 in $H^0(B^{\otimes 4})$.
- (b) If $r = 1$ and $n > 2$, γ_l surjects for all $l \geq 1$, except if $l = 1, 3$. The image of γ_1 has codimension $n - 2$ in $H^0(B^{\otimes 2})$. The map $\gamma(2, 2)$ is surjective.
- (c) If $r \geq 2$, γ_l is surjective if $l \geq 3$. The image of γ_1 has codimension $r(n - 2)$ in $H^0(B^{\otimes 2})$. The image of γ_2 has codimension $r - 1$ in $H^0(B^{\otimes 3})$.

Now a suitable hyperplane section of the image of φ is an irreducible surface Y of minimal degree. Its pullback by φ to X is a surface S of general type with at worst canonical singularities. Moreover, by adjunction $B \otimes \mathcal{O}_S = K_S$, and the complete linear series of B restricts to the complete canonical linear series of S ; so $\varphi|_S$ is the canonical morphism of S . Therefore S is under the hypothesis of Theorem 2.1, and the proof follows verbatim the steps given in 2.7, the role of S there being played by X here, the role of C there being played by S here and the role of Proposition 2.2 there being played by Theorem 2.1 here.

To show what we mean, we outline how to find the images of the maps γ_l . We consider the following commutative diagram:

$$\begin{array}{ccccc} H^0(B^{\otimes l}) \otimes H^0(\mathcal{O}_X) & \hookrightarrow & H^0(B^{\otimes l}) \otimes H^0(B) & \twoheadrightarrow & H^0(B^{\otimes l}) \otimes H^0(K_S) \\ \downarrow & & \downarrow \gamma_l & & \downarrow \\ H^0(B^{\otimes l}) & \hookrightarrow & H^0(B^{\otimes l+1}) & \twoheadrightarrow & H^0(K_S^{\otimes l+1}) \end{array}$$

The rightmost horizontal arrows are surjective because $H^1(\mathcal{O}_X) = 0$, by Serre duality and by Kawamata-Viehweg vanishing. The left vertical arrow trivially surjects. The right vertical arrow is the composition of the map $H^0(B^{\otimes l}) \otimes H^0(K_S) \rightarrow H^0(K_S^{\otimes l}) \otimes H^0(K_S)$, which is surjective for all $l \geq 1$ again because $H^1(\mathcal{O}_X) = 0$, by Serre duality and by Kawamata-Viehweg vanishing, and the map α_l of multiplication of global sections on S , studied in (2.7). Then it follows from chasing the diagram that the map $H^0(B^{\otimes l+1}) \rightarrow H^0(K_S^{\otimes l+1})$ maps the image of γ_l onto the image of α_l and that the codimension of the image of α_l in $H^0(K_S^{\otimes l+1})$ (which was equal to the codimension of the image of β_l in $H^0(\theta^{\otimes l+1})$) is equal to the codimension of γ_l in $H^0(B^{\otimes l+1})$. This, together with the claims in (a), (b) and (c) concerning the codimensions of the images of the maps α_l proved in (2.7), gives us the codimension of the images of the maps γ_l in $H^0(B^{\otimes l+1})$. The claims regarding $\gamma(2, 2)$ are proved analogously. \square

As we did in the case of the canonical ring of regular surfaces of general type, we can characterize when $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is generated in degree less than or equal to 2 using Theorem 4.1 and results from [GP2]:

Corollary 4.2. *Let X be a Calabi-Yau threefold with at worst canonical singularities. Let B be a big and base-point-free line bundle on X and let φ be the morphism induced by $|B|$. Then $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is generated in degree less than or equal to 2 if and only if*

- (1) φ does not map X generically 2 : 1 onto \mathbf{P}^3 ; and
- (2) $\varphi(X)$ is not a threefold of minimal degree other than \mathbf{P}^3 .

Proof. Theorem 4.1 tells us that if $\varphi(X)$ is a variety of minimal degree, then the ring $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is generated in degree less than or equal to 2 if and only if $\varphi(X) = \mathbf{P}^3$ and the degree of φ is greater than 2. If $\varphi(X)$ is not a variety of minimal degree, then in the proofs of [GP2], Theorems 1.4 and 1.7, it is shown that

$$H^0(B^{\otimes l}) \otimes H^0(B^{\otimes 1}) \xrightarrow{\gamma_l} H^0(B^{\otimes l+1})$$

surjects if $l \geq 2$. \square

The study of the generators of the ring $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is closely related with the question of when $B^{\otimes m}$ is normally generated when B is ample. Recall that a line bundle L is said to be normally generated, or to satisfy property N_0 , if it is very ample and the image of the morphism induced by $|L|$ is a projectively normal variety. This is equivalent to the ring $\bigoplus_{k=0}^{\infty} H^0(L^{\otimes k})$ being generated in degree 1. In the present context (X a Calabi-Yau threefold and B an ample and base-point-free line bundle on X), the answer to the question of when $B^{\otimes m}$ is normally generated is partially known. If the image of X by the morphism φ induced by $|B|$ is not a variety of minimal degree, the authors gave a complete answer to this question for $m \geq 2$: in this case, in [GP2], Corollary 1.1, Theorem 1.4 and Theorem

1.7, they proved that $B^{\otimes m}$ is normally generated if $m \geq 2$. The way of proving those results serves to illustrate the relation between the study of the generators of $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ and the normal generation of powers of B . For instance, in order to prove the normal generation of $B^{\otimes 2}$, the first step is to show that the map

$$H^0(B^{\otimes 2}) \otimes H^0(B^{\otimes 2}) \longrightarrow H^0(B^{\otimes 4})$$

is surjective. This was proved in [GP2] by showing that $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is generated in degree less than or equal to 2 and that the map

$$H^0(B^{\otimes 3}) \otimes H^0(B) \longrightarrow H^0(B^{\otimes 4})$$

is surjective. If $\varphi(X)$ is a variety of minimal degree, the situation is more complex. If $m \geq 3$, the authors also gave in [GP2], Corollary 1.1 and Theorem 1.4, a complete answer: $B^{\otimes m}$ is normally generated if and only if $m \geq 4$ or if $m \geq 3$ and φ does not map X 2 : 1 onto linear \mathbf{P}^3 . If $m = 2$, answering the question will settle the following

Conjecture 4.3 (cf. [GP2], Conjecture 1.9). *Let X be a Calabi-Yau threefold and let B be an ample and base-point-free line bundle. Then $B^{\otimes 2}$ satisfies property N_0 if and only if there is a smooth non-hyperelliptic curve C in $|B \otimes \mathcal{O}_S|$, for some $S \in |B|$.*

This conjecture would also give a characterization of when $B^{\otimes 2}$ is very ample. This question of when $B^{\otimes 2}$ is very ample is also open.

One might ask what light the results proved in this section shed on the conjecture. Theorem 4.1 says that if $\varphi(X)$ is \mathbf{P}^3 , then the ring $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is generated in degree less than or equal to 2 if and only if the degree of φ is greater than 2. In this case, however, it was seen in the proof of Theorem 4.1 that γ_3 did not surject. On the other hand, if $\varphi(X)$ is a variety of minimal degree different from \mathbf{P}^3 , γ_3 does surject but $\bigoplus_{l=0}^{\infty} H^0(B^{\otimes l})$ is not generated in degree less than or equal to 2. Therefore the strategy outlined before to study the normal generation of $B^{\otimes 2}$, which worked when $\varphi(X)$ was not of minimal degree, does not work if $\varphi(X)$ is of minimal degree. We point out that the conjecture is nevertheless true if $\varphi(X)$ is linear \mathbf{P}^3 (see [GP2], Corollary 1.8). This also follows from the methods of this article, by studying the map $\gamma(2, 2)$ in the proof of Theorem 4.1. Thus the only case left in order to settle the conjecture is when $\varphi(X)$ is a variety of minimal degree different from \mathbf{P}^3 , which should be addressed using a subtler strategy.

The results of Section 3 regarding the structure of the canonical morphisms of regular surfaces of general type onto surfaces of minimal degree has some interesting consequences for Calabi-Yau threefolds. As we will see, Proposition 3.1, Corollary 3.2 and Proposition 3.3 prevent many a priori natural examples of Calabi-Yau threefolds from existing. This also shows that if there do exist examples of prime degree Calabi-Yau covers of threefolds of minimal degree induced by complete linear series, then they cannot come from group actions. We summarize this in the next two corollaries:

Corollary 4.4. *Let X be a Calabi-Yau threefold with at worst canonical singularities. Let B be a base-point-free line bundle. Let $X \xrightarrow{\varphi} Y$ be the morphism induced by the complete linear series $|B|$. Let n be the degree of φ . Let U be a smooth open set of Y whose complement has codimension 2, and let L be a line bundle on U . Assume that*

$$(*) \quad (\varphi_* \mathcal{O}_X)|_U = \mathcal{O}_U \oplus L^{-1} \oplus \dots \oplus L^{\otimes 1-n}.$$

If Y is a variety of minimal degree, then the degree $n = 2$ or $n = 3$. In particular, if φ is a cyclic cover, the degree of φ is 2 or 3

Proof. Let Y' be a suitable hyperplane section of Y , and S the pullback of Y' by φ . Then $\varphi|_S$ is the canonical morphism of S and satisfies the hypothesis of Proposition 3.1, and the thesis is clear. \square

Notation 4.5. We will call a morphism φ satisfying (*) in the statement of Corollary 4.4 a quasi-cyclic cover.

Corollary 4.6. *Let X be a Calabi-Yau threefold with at worst canonical singularities. If $X \xrightarrow{\varphi} Y$ is a generically finite morphism onto a smooth scroll $Y \subset \mathbf{P}^N$, then the degree of φ is even. In particular, there are no generically finite covers of odd degree of smooth rational normal scrolls.*

Proof. The proof is analogous to that of Corollary 4.4, using now Proposition 3.3. \square

In [GP2] we described what finite morphisms from a Calabi-Yau threefold onto a variety of minimal degree induced by complete linear series were possible. After Corollary 4.4 and Corollary 4.6 we can obtain the following sharper version of the result in [GP2] (compare with [GP2], Proposition 1.6):

Proposition 4.7. *Let X be a smooth Calabi-Yau threefold, let φ be the morphism induced by the complete linear series of an ample and base-point-free line bundle B on X , and let n be the degree of φ . If $\varphi(X)$ is a variety of minimal degree, then one of the following occurs:*

- 1) $Y = \mathbf{P}^3$ and $n \leq 24$.
- 2) Y is a smooth quadric hypersurface in \mathbf{P}^4 and $n = 2, 4, 6, 8, 10, 12$ or 14.
- 3) Y is a smooth rational normal scroll of dimension 3 in \mathbf{P}^5 and $n = 2, 4, 6, 8, 10$ or 12.
- 4) Y is a smooth rational normal scroll of dimension 3 in \mathbf{P}^N , $N \geq 6$, $n = 2$, and X is fibered over \mathbf{P}^1 with a smooth K3 surface as a general fiber. The restriction of B to the general fiber of X is hyperelliptic, with sectional genus 2, and its complete linear series maps the fiber 2 : 1 onto a general fiber of the scroll.
- 5) Y is a smooth rational normal scroll of dimension 3 in \mathbf{P}^N , $N \geq 6$, $n = 6$, and X is fibered over \mathbf{P}^1 with a smooth abelian surface as a general fiber. The restriction of B to the general fiber of X is a (1, 3) polarization, and its complete linear series maps the fiber 6 : 1 onto a general fiber of the scroll.
- 6) Y is a cone over a conic in \mathbf{P}^2 .
- 7) Y is a cone over a twisted cubic in \mathbf{P}^3 .
- 8) Y is a cone over a Veronese surface.

In addition, if X has at worst canonical singularities and φ is a quasi-cyclic cover, then $n = 2$ or 3.

REFERENCES

- [Bo] E. Bombieri, *Canonical models of surfaces of general type*, Inst. Hautes Etudes Sci. Publ. Math. **42** (1973), 171–219. MR **47**:6710
- [Ca] F. Catanese, *On the moduli spaces of surfaces of general type*, J. Differential Geometry **19** (1984), 483–515. MR **86h**:14031
- [Ci] C. Ciliberto, *Sul grado dei generatori dell'anello di una superficie di tipo generale*, Rend. Sem. Mat. Univ. Politec. Torino **41** (1983), 83–111. MR **86d**:14036

- [EH] D. Eisenbud and J. Harris, *On varieties of minimal degree (a centennial account)*, Algebraic Geometry, Bowdoin 1985, Amer. Math. Soc. Sympos. in Pure and Appl. Math. **46** (1987), 1–14. MR **89f**:14042
- [GP1] F. J. Gallego and B. P. Purnaprajna, *Projective normality and syzygies of algebraic surfaces*, J. Reine Angew. Math. **506** (1999), 145–180. MR **2000a**:13023; MR **2001b**:13016
- [GP2] F. J. Gallego and B. P. Purnaprajna, *Very ampleness and higher syzygies for Calabi-Yau threefolds*, Math. Ann. **312** (1998), 133–149. MR **99g**:14048
- [GP3] F. J. Gallego and B. P. Purnaprajna, *Canonical covers of varieties of minimal degree*, Preprint math.AG/0205010. To appear in “A tribute to Seshadri—a collection of papers on Geometry and Representation Theory”, Hindustan Book Agency (India) Ltd.
- [GP4] F. J. Gallego and B. P. Purnaprajna, *Some homogeneous rings associated to finite morphisms*, Preprint. To appear in “Advances in Algebra and Geometry” (Hyderabad Conference 2001), Hindustan Book Agency (India) Ltd.
- [GP5] F. J. Gallego and B. P. Purnaprajna, *On the rings of trigonal curves*, in preparation.
- [G] M. L. Green, *The canonical ring of a variety of general type*, Duke Math. J. **49** (1982), 1087–1113. MR **84k**:14006
- [HM] D. Hahn and R. Miranda, *Quadruple covers of algebraic varieties*, J. Algebraic Geom. **8** (1999), 1–30. MR **99k**:14028
- [H1] E. Horikawa, *Algebraic surfaces of general type with small c_1^2 . I*, Ann. of Math. (2) **104** (1976), 357–387. MR **54**:12789
- [H2] E. Horikawa, *Algebraic surfaces of general type with small c_1^2 , II*, Invent. Math. **37** (1976), 121–155. MR **57**:334
- [H3] E. Horikawa, *Algebraic surfaces of general type with small c_1^2 , III*, Invent. Math. **47** (1978), 209–248. MR **80h**:14012a
- [H4] E. Horikawa, *Algebraic surfaces of general type with small c_1^2 , IV*, Invent. Math. **50** (1978/79), 103–128. MR **80h**:14012b
- [Kod] K. Kodaira, *Pluricanonical systems on algebraic surfaces of general type*, J. Math. Soc. Japan **20** (1968), 170–192. MR **37**:212
- [Kon] K. Konno, *Algebraic surfaces of general type with $c_1^2 = 3p_g - 6$* , Math. Ann. **290** (1991), 77–107. MR **92i**:14039
- [MP] M. Mendes Lopes and R. Pardini, *Triple canonical surfaces of minimal degree*, International J. Math. **11** (2000), 553–578. MR **2001h**:14049
- [M] D. Mumford, *Varieties defined by quadratic equations*, Corso CIME in Questions on Algebraic Varieties, Edizioni Cremonese, Rome (1970), 29–100. MR **44**:209
- [OP] K. Oguiso and T. Peternell, *On polarized canonical Calabi-Yau threefolds*, Math. Ann. **301** (1995), 237–248. MR **96b**:14050
- [R] M. Reid, *Infinitesimal view of extending a hyperplane section—deformation theory and computer algebra*, Algebraic geometry, Proceedings L'Aquila 1988, 214–286, Lecture Notes in Math. **1417**, Springer-Verlag, Berlin, 1990. MR **91h**:14018

DEPARTAMENTO DE ÁLGEBRA, FACULTAD DE CIENCIAS MATEMÁTICAS, UNIVERSIDAD COMPLUTENSE DE MADRID, 28040 MADRID, SPAIN

E-mail address: FJavier.Gallego@mat.ucm.es

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF KANSAS, 405 SNOW HALL, LAWRENCE, KANSAS 66045-2142

E-mail address: purna@math.ukans.edu

A CLASSIFICATION AND EXAMPLES OF RANK ONE CHAIN DOMAINS

H. H. BRUNGS AND N. I. DUBROVIN

ABSTRACT. A chain order of a skew field D is a subring R of D so that $d \in D \setminus R$ implies $d^{-1} \in R$. Such a ring R has rank one if $J(R)$, the Jacobson radical of R , is its only nonzero completely prime ideal. We show that a rank one chain order of D is either invariant, in which case R corresponds to a real-valued valuation of D , or R is nearly simple, in which case R , $J(R)$ and (0) are the only ideals of R , or R is exceptional in which case R contains a prime ideal Q that is not completely prime. We use the group $\mathcal{M}(R)$ of divisorial R -ideals of D with the subgroup $\mathcal{H}(R)$ of principal R -ideals to characterize these cases. The exceptional case subdivides further into infinitely many cases depending on the index k of $\mathcal{H}(R)$ in $\mathcal{M}(R)$. Using the covering group \mathbb{G} of $\mathrm{SL}(2, \mathbb{R})$ and the result that the group ring $T\mathbb{G}$ is embeddable into a skew field for T a skew field, examples of rank one chain orders are constructed for each possible exceptional case.

INTRODUCTION

A subring R of a skew field D is called total if d in D and d not in R implies that the inverse d^{-1} is contained in R . It follows that for such rings R the lattice of right ideals as well as the lattice of left ideals is linearly ordered by inclusion; R is a chain domain. Conversely, any chain domain R is Ore and is a total subring of its skew field of quotients D . The total subrings of fields are exactly valuation rings, corresponding to valuation functions into linearly ordered groups. In particular, if we take nontrivial subgroups G of the additive group $(\mathbb{R}, +, \leq)$ of the reals as value groups, then we obtain the commutative valuation rings of rank one. Such a ring can also be characterized as a maximal subring of a field, or as a valuation ring with exactly one nonzero prime ideal. In the non-commutative case we must distinguish between prime ideals and completely prime ideals: An ideal $B \neq R$ of a ring R is prime if $I_1 I_2 \subseteq B$ implies $I_1 \subseteq B$ or $I_2 \subseteq B$ for ideals I_1 and I_2 of R . If $ab \in B$ implies $a \in B$ or $b \in B$ for elements a, b in R , then B is called completely prime. A total subring R of a skew field D will be called a chain domain of rank one if R has exactly one nonzero completely prime ideal. This ideal will then be $J(R)$, the Jacobson radical of R .

Received by the editors April 10, 2002 and, in revised form, October 9, 2002.

2000 *Mathematics Subject Classification.* Primary 16L30, 16K40, 16W60; Secondary 20F29, 20F60.

Key words and phrases. Exceptional chain domains, skew field, valuation, cone, covering group. The first author was supported by NSERC.

The second author was supported by RFBR and DFG (grant no. 98-01-04110).

We prove in Theorem 1.9 that a rank one chain domain R is either invariant, i.e., all one-sided ideals are two-sided, or it is nearly simple in which case R , $J(R)$, and (0) are its only ideals, or R is exceptional in which case R contains a prime ideal that is not completely prime. The exceptional rank one chain domains are classified further with the help of the group $\mathcal{M}(R)$ of divisorial R -ideals and the subgroup $\mathcal{H}(R)$ of $\mathcal{M}(R)$ of principal R -ideals. The lattice of two-sided R -ideals is then determined by the index k of $\mathcal{H}(R)$ in $\mathcal{M}(R)$, and we say that R is exceptional of type (C_k) .

These results are proved in the more general case of cones P in groups G where a cone P of G is a subsemigroup of G so that $g \in G \setminus P$ implies $g^{-1} \in P$.

That rank one chain domains are either invariant, nearly simple or exceptional was proved in [4]. Invariant rank one chain orders of D correspond to valuation functions from D^* into $(\mathbb{R}, +, \leq)$. Nearly simple chain domains were constructed in [8], [16], [5] and [3]. The construction of exceptional rank one chain domains, however, appeared to be elusive even though Posner in [19] hinted that such rings might exist, and the classification of hypercyclic rings by Osofsky in [18] is complete only if such rings do not exist. I. N. Herstein had considered the problem and this existence problem was also encountered in [14]. We construct in this paper exceptional rank one chain domains of any type (C_k) : Theorem 4.4 and Corollary 4.6. We do this by first constructing exceptional cones P_k of type (C_k) in subgroups H_k of the universal covering group \mathbb{G} of $\mathrm{SL}(2, \mathbb{R})$, Theorem 3.8, and then apply Dubrovin's result in [11], where he constructs an exceptional rank one chain ring of type (C_1) associated with a cone \mathbb{P} in \mathbb{G} .

1. CHAIN DOMAINS AND CONES

1.1. Basic properties. A ring R is a *right chain ring*, if the set of all right ideals of R is linearly ordered with respect to inclusion. Left chain rings and chain rings are defined similarly. A chain domain R has a classical skew field of quotients D and can therefore be considered as a total subring of D ([7]).

A subsemigroup P of a group G is called a *cone* of G if $G = P \cup P^{-1}$ and P is a *pure cone* if in addition $P \cap P^{-1} = \{e\}$. There is a close connection between cones P in a group G and right or left orders: if P is a cone of G and $a, b \in G$, then \leq_ℓ defined by $a \leq_\ell b$ if and only if $a^{-1}b \in P$ defines a left preorder, and $a \leq_r b$ if and only if $ba^{-1} \in P$ defines a right preorder on G . The relations " \leq_r " and " \leq_ℓ " are right orders and left orders on G respectively if and only if the cone P is pure. Finally, if P is pure, then the right order defined by P agrees with the left order defined by P if and only if $aP = Pa$ for all a in G , i.e., P is invariant under all inner automorphisms of G . The group G is then linearly ordered.

Let P be a cone of a group G . A nonempty subset I of G is called a *left P -ideal* if $PI \subseteq I$ and $I \subseteq Pa$ for a suitable element a in G . The second condition is satisfied for any $I \neq G$ provided I satisfies the first condition. If in addition $I \subseteq P$, we say I is a *left ideal*. Right P -ideals, P -ideals and right ideals and ideals are defined similarly. An ideal B of P is called a *prime ideal* if $B \neq P$ and $aPb \subseteq B$ implies $a \in B$ or $b \in B$ for $a, b \in P$. If $ab \in B$ implies $a \in B$ or $b \in B$ for the ideal $B \neq P$ of P , then B is called *completely prime*.

We collect elementary properties of a cone P in G . We can assume that $P \neq G$. Let $U(P) = P \cap P^{-1}$, the subgroup of units of P .

- a): $J(P) = P \setminus U(P)$ is the maximal right and the maximal left ideal of P ; it is the Jacobson radical of P and it is a completely prime ideal of P .
- b): The set of right (left) P -ideals in G is linearly ordered with respect to inclusion. We define $I_1 \leq I_2$ if and only if $I_1 \supseteq I_2$ for right P -ideals I_1 and I_2 .

To see this, one considers first principal right P -ideals aP and bP in G . Then either $a^{-1}b \in P$ and $bP \subseteq aP$ or $b^{-1}a \in P$ and $aP \subseteq bP$. If $I_2 \not\subseteq I_1$, then there exists a in $I_2 \setminus I_1$ and $I_1 \subset aP \subseteq I_2$ follows.

- c): There is a one-to-one correspondence between the set of cones $P' \neq G$ in G that contain a cone P and the set of completely prime ideals B of P .

Proof. Let $P \subseteq P' \subset G$ be cones in G . Then $j' \in J(P')$ and $j' \notin P$ implies $j'^{-1} \in P$, a contradiction. Hence, $J(P') \subseteq J(P)$ and $P' = P \cup (P \setminus J(P'))^{-1}$. Conversely, if $B \subseteq J(P)$ is a completely prime ideal in P , then $P' = P \cup (P \setminus B)^{-1}$ is a cone $\neq G$ in G . \square

- d): Let I be an ideal in P with $I \neq P$ and $Q = \bigcap I^n \neq \emptyset$. Then Q is a completely prime ideal.

Proof. If $c \in P \setminus Q$ and $ca \in Q$ for some a in P , then there exists n_0 with $c \notin I^{n_0}$. However, for any n there exist $a_i, b_j \in I$ with $ca = a_1 \dots a_{n_0} b_1 \dots b_n$. Then $a_1 \dots a_{n_0} = cd$ for some d in P and $a = db_1 \dots b_n \in I^n$ follows. Hence, $a \in Q$ and Q is a completely prime ideal. \square

- e): A P -ideal I will be right principal and left principal if and only if $I = {}_zP = Pz$ for some $z \in G$.

Proof. Let $I = z_1P = Pz_2$ with $z_1, z_2 \in G$. Then $z_2 = z_1a$, $z_1 = bz_2$ for some $a, b \in P$. Hence, $bz_1a = z_1$. Since I is an ideal, there exists b' in P with $bz_1 = z_1b'$ and $z_1 = z_1b'a$ follows. Therefore, $b'a = 1$ and $a \in U(P)$, and $Pz_2 = z_1P = z_1aP = z_2P$. \square

- f): Let P be a cone in G . The set $\mathcal{H}(P)$ of all principal P -ideals of G forms a group with ideal multiplication as the operation. $\mathcal{H}(P)$ is isomorphic to a subgroup of $(\mathbb{R}, +, \leq)$ if $J(P)$ is the only completely prime ideal of P .

Proof. If $I_1 = z_1P = Pz_1$ and $I_2 = z_2P = Pz_2$, then $I_1I_2 = z_1Pz_2P = z_1z_2P$ and $(z_1P)^{-1} = z_1^{-1}P$. It follows that $\mathcal{H}(P)$ is a group with P as identity. To prove the second statement let $P \supset {}_zP = Pz$. Then $\bigcap ({}_zP)^n = \emptyset$ since otherwise $\bigcap ({}_zP)^n$ is a completely prime ideal $\neq J(P)$ by d). $\mathcal{H}(P)$ is therefore an ordered Archimedean group and the statement follows from Hölder's Theorem (see [13]). \square

- g): A right P -ideal I is a principal right P -ideal if and only if $IJ(P) \neq I$.

Proof. If $I = {}_zP$, then ${}_zP \supset IJ(P) = {}_zJ(P)$. Conversely, if I is not principal as a right P -ideal, then for $a \in I$ there exists $b \in I$ with $aP \subset bP$, $a = bj \in IJ(P)$, $j \in J(P)$, and $IJ(P) = I$. \square

We single out cones with the property in f):

Definition 1.1. A cone P of a group G has *rank one* if $J(P)$ is the only completely prime ideal of P .

It follows from the definitions that a subring R of a skew field D is total if and only if the semigroup $R^* = (R \setminus \{0\}, \cdot)$ is a cone in the group D^* .

This relationship between a cone in a group and a chain domain is generalized in the next definition.

Definition 1.2. A total subring R in a skew field D is said to be *associated* with a cone P in a group G if the following conditions hold:

- i): G is a subgroup of D^* , the multiplicative group of D .
- ii): Every element d in D^* can be written as $d = g_1 u_1 = u_2 g_2$ with g_1, g_2 in G and u_1, u_2 in $U(R)$ so that $P g_1 P = P g_2 P$.
- iii): $R \cap G = P$.

We also say in this case that the cone P is associated with the chain domain R .

Proposition 1.3. Let the total subring R of the skew field D be associated with the cone P of the group G . Then:

- i): $I_0 \rightarrow I_0 R$ defines an isomorphism from the lattice of right P -ideals to the lattice of nonzero right R -ideals. The inverse of this mapping assigns $I \cap G$ to the nonzero right R -ideal I .
- ii): The correspondence defined in i) preserves the properties of being an ideal, a completely prime ideal, a prime ideal, and a principal right ideal.

Proof. i) If I_0 is a right P -ideal, then two nonzero elements a, b in $I_0 R$ have the form $a = g_1 u_1, b = g_2 u_2$ for $g_i \in I_0$ and $u_i \in U(R)$. We can assume that $g_1 P \subseteq g_2 P$, and $g_1 = g_2 p, p \in P$ follows. Therefore, $a \pm b = g_2 (p u_1 \pm u_2) \in I_0 R$; this shows that $I_0 R$ is a right R -ideal, since $g I_0 \subseteq P \subseteq R$ for some $g \in G \subseteq D$. Further, if $g \in I_0 R \cap G$ for a right P -ideal I_0 , then $g = h g' u$ for $h \in I_0, g' \in P$ and $u \in U(R)$. It follows that $h g' \in I_0$ and $u \in U(R) \cap G = U(P)$; hence, $g \in I_0$ and $I_0 R \cap G = I_0$. Similarly, one can show that $I \cap G$ is a right P -ideal if I is a right R -ideal and that $(I \cap G)R = I$.

For ii) we only show that the right P -ideal I_0 is a P -ideal if and only if $I_0 R$ is an R -ideal. Let $r \in R$ and $h \in I_0$, a P -ideal. Then $r = p_1 u_1$ for $p_1 \in P, u_1 \in U(R)$ and $r h = p_1 u_1 h = p_1 k u_2$ for $u_1 h = k u_2$ with $u_2 \in U(R)$ and $h, k \in G$. By ii) of Definition 1.2 we have $P h P = P k P$; $k \in I_0$ follows and $r h \in I_0 R$, which shows that $I_0 R$ is also a left R -module and then an R -ideal. Conversely, if $I_0 R$ is an R -ideal for a right P -ideal I_0 , then $I_0 = I_0 R \cap G$ is a P -ideal. \square

Some variations of the results in this section can be found in [12] and [6].

1.2. Divisorial ideals. We consider certain P -ideals for a cone P which will form a group in case P has rank one.

Definition 1.4. Let P be a cone in a group G . The *divisorial closure* \widehat{I} of a right P -ideal I is the intersection of all principal right P -ideals containing I :

$$\widehat{I} = \bigcap_{hP \supseteq I} hP.$$

A right P -ideal I is called *divisorial* if $\widehat{I} = I$.

If we replace the cone P by a total subring R , we obtain the definition of the *divisorial closure* of a right R -ideal and of a *divisorial* right R -ideal. In addition, we assume that a divisorial right R -ideal is nonzero.

We collect a list of properties:

Let P be a cone in a group G , I a P -right ideal. Then:

a): $\widehat{I} \supseteq I$;

b): $\widehat{\widehat{I}} = \widehat{I}$;

c): $g\widehat{I} = g\widehat{I}$ for any g in G ;

d): I is non-divisorial if and only if $J(P)$ is not a principal right ideal and there exists an element z in G with $\widehat{I} = zP$ and $I = zJ(P)$. If, in addition, I is a P -ideal and $\text{rank } P = 1$, then $\widehat{I} = zP = Pz$ and $I = zJ(P) = J(P)z$.

The properties a, b, and c follow directly from the definition. To prove d) we will write J instead of $J(P)$ and assume that $\widehat{I} \supset I$ and that $z \in \widehat{I} \setminus I$. Then $\widehat{I} \supseteq zP \supset I$ and $\widehat{I} = zP$ follows; then $I = zJ$, since $zjP \supseteq I$ for some $j \in J(P)$ leads to a contradiction. This also shows that J is not a principal right ideal. If J is not a principal right ideal, then $cP \supseteq zJ$ implies $z^{-1}cP \supseteq J$ and $z^{-1}cP \supseteq P$, $cP \supseteq zP$ for $c, z \in G$. This means that $\widehat{I} = zP$ for $I = zJ$ and hence $\widehat{I} \supset I$. If zP is a P -ideal, then certainly zJ is a P -ideal. Conversely, if zJ is an ideal, then zP is an ideal, since otherwise there is an $a \in P$ and a $j \in J$ with $azj = z$, a contradiction. Finally, we assume that $\widehat{I} \neq I$ and I is a P -ideal and that P has rank one. Then $\widehat{I} = zP$ and the left order $O_\ell(I) = \{g \in G \mid g\widehat{I} \subseteq \widehat{I}\} \neq G$ contains the cone P as well as the cone zPz^{-1} both of which are maximal. It follows that $P = zPz^{-1}$, $Pz = zP = \widehat{I}$ and $Jz = zJ = I$. \square

We list a property that was proved in the proof of d):

e): I is a P -ideal if and only if \widehat{I} is a P -ideal.

The next result shows that in the correspondence between right R -ideals and right P -ideals, divisorial right ideals correspond to each other if the chain domain R is associated with the cone P .

Proposition 1.5. *Let R be a total subring of the skew field D associated with the cone P in a group G . Then the right P -ideal I is divisorial if and only if the right R -ideal IR is divisorial.*

Proof. Assume I is divisorial, i.e., $I = \widehat{I} = \bigcap_{hP \supseteq I} hP$. Then $IR \subseteq hR$ for all $h \in G$

with $hP \supseteq I$, and $IR \subseteq \bigcap_{hP \supseteq I} hR$. To show the reverse inclusion, let $d \in \bigcap_{hP \supseteq I} hR$

for $hP \supseteq I$ and $d = hr_h = gm$ for $g \in G$, $m \in U(R)$. Hence, $g = hr_h m^{-1} \in hR \cap G = hP$ and $g \in \bigcap_{hP \supseteq I} hP = I$, $d \in IR$ follows. Now assume that A is a divisorial right R -ideal, $A = \bigcap_{dR \supseteq A} dR$. Any such $d = gm$ for $g \in G$, $m \in U(R)$. Hence,

$A \cap G = (\bigcap_{dR \supseteq A} dR) \cap G = \bigcap (gR \cap G) = \bigcap gP$, which shows that $A \cap G$ is divisorial and $A \cap G$ is nonempty, since A is nonzero. \square

For any subset I of a group G we define the following three subsets of G : the right order $O_r(I) = \{g \in G \mid Ig \subseteq I\}$, the left order $O_\ell(I) = \{g \in G \mid gI \subseteq I\}$, and the inverse $I^{-1} = \{g \in G \mid Ig \subseteq I\}$.

It follows that $I^{-1} = \{g \in G \mid gI \subseteq O_r(I)\} = \{g \in G \mid Ig \subseteq O_\ell(I)\}$.

We have the following two properties where P is a cone in the group G :

f): If I is a right P -ideal, then $O_\ell(I)$ is a cone of G and $O_r(I)$ is an over cone of P . Further, I is a right $O_r(I)$ -ideal and a left $O_\ell(I)$ -ideal, and I^{-1} is a right $O_\ell(I)$ -ideal and a left $O_r(I)$ -ideal.

For a proof we observe that for any g in G either $gI \subseteq I$ and $g \in O_\ell(I)$ or $I \subset gI$ and $g^{-1} \in O_\ell(I)$. The rest of the statements follow immediately.

g): $O_r(J(P)) = O_\ell(J(P)) = P$, and $J(P)^2 \neq J(P)$ implies that $J(P) = {}_zP = Pz$ for some $z \in P$.

The first statement follows from Property c) in Section 1.1 since $O_r(J(P)) \supset P$ implies that $j^{-1}J(P) \subseteq J(P)$ for some $j \in J(P)$. Hence, $J(P) \subseteq jJ(P)$, a contradiction that shows $O_r(J(P)) = P$ and similarly $O_\ell(J(P)) = P$.

The second statement follows from Property g) in Section 1.1, its left symmetric version, and Property e) in Section 1.1. \square

Even though one can consider the groupoid of all divisorial P -ideals for a cone P of arbitrary rank (see also [2]), we restrict ourselves to the rank one case:

Definition 1.6. Let P be a cone of rank one. Then $\mathcal{M}(P)$ is the set of all divisorial P -ideals together with the operation “ $*$ ” defined by:

$$I_1 * I_2 = \widehat{I_1 I_2} \quad \text{for } P\text{-ideals } I_1, I_2.$$

We have the following result:

Theorem 1.7. Let P be a cone of rank one in a group G . Then:

- $\alpha)$ $\mathcal{M}(P)$ is a linearly ordered group;
- $\beta)$ The inverse of an element I in $\mathcal{M}(P)$ is I^{-1} ;
- $\gamma)$ $\mathcal{H}(P)$ is a subgroup of $\mathcal{M}(P)$.

Proof. We show first that the operation defined in Definition 1.6 is associative.

On the set of all P -ideals we define a relation $I_1 \sim I_2$ if and only if $\widehat{I_1} = \widehat{I_2}$; this is an equivalence relation.

We are going to show next that for P -ideals I_1, I_2 the following equivalence holds:

$$(+) \quad I_1 I_2 \sim \widehat{I_1} \widehat{I_2}.$$

If $I_1 = \widehat{I_1}$ and $I_2 = \widehat{I_2}$, then $(+)$ is trivially true. If $I_1 \neq \widehat{I_1}$, then $\widehat{I_1} = {}_zP \supset {}_zJ(P) = I_1$ and $J = J(P)$ is not right principal. Also $\widehat{I_1} = {}_zP = Pz$ is a P -ideal by Property d).

The equivalence $(+)$ holds therefore if and only if the following equivalence holds:

$$(++) \quad JI_2 \sim P\widehat{I_2} = \widehat{I_2}.$$

Hence, if $JI_2 = I_2$, we are done. Otherwise, $JI_2 \subset I_2$ and $I_2 = Pd$ follows for some d in G by the left symmetric version of Property g) in Section 1.1. Since I_2 is an ideal, we have $dP \subseteq Pd$, $P \subseteq d^{-1}Pd$ and the equality $d^{-1}Pd = P$ since P has rank one. Then $dP = Pd = I_2$, $dJ = Jd$ and $JI_2 = Jd = dJ \sim dP = I_2$ which proves the equivalence $(++)$ and hence also $(+)$ in this case.

Finally, we must prove $(+)$ if $I_1 = \widehat{I_1}$ and $\widehat{I_2} \supset I_2$. Then, as above, $\widehat{I_2} = aP = Pa \supset aJ = Ja = I_2$ for some a in G . The equivalence $(+)$ then holds if and only if the equivalence $I_1 J \sim \widehat{I_1} P = I_1$ holds. Using the right symmetric version of arguments used in the proof of $(++)$, one shows that $I_1 J \sim I_1$. This proves $(+)$.

If $I_1 \sim I'_1$ and $I_2 \sim I'_2$ for P -ideals I_1, I'_1, I_2, I'_2 , then $I_1 I_2 \sim \widehat{I_1} \widehat{I_2} = \widehat{I'_1} \widehat{I'_2} \sim I'_1 I'_2$. Hence $\mathcal{M}(P)$ is a factor monoid of the monoid of all P -ideals, and the operation $*$ given in the definition for $\mathcal{M}(P)$ is associative.

Next we show that $\widehat{II^{-1}} = P$ for I a P -ideal, and $\widehat{I^{-1}I} = P$ follows from similar arguments. Since I is a P -ideal, I^{-1} is a P -ideal.

If $II^{-1} = P$, we are done; otherwise $II^{-1} \subseteq J(P) = J$. If $II^{-1} \subseteq Pz \subseteq J$ for some $z \in J$, then $II^{-1}z^{-1} \subseteq P$ and $I^{-1}z^{-1} \in I^{-1}$ which implies $z^{-1} \in O_r(I^{-1}) = P$, since P has rank one. This is a contradiction since $z \in J$, and $II^{-1} = J$, $J \neq Pz$ for all $z \in P$ remains as the only possibility to be considered. It then follows from Property g) that $J^2 = J$, J is not a principal right ideal, and hence $\widehat{II^{-1}} = \widehat{J} = P$.

In order to complete the proof of $\alpha)$ and $\beta)$ we show that I^{-1} is a divisorial P -ideal for I a P -ideal. If on the contrary, $I^{-1} = zJ \subset zP = \widehat{I^{-1}}$ and J is not a principal right ideal, then $zJI \subseteq P$ by the definition of I^{-1} , and by (+) it follows that $z\widehat{JI} \subseteq \widehat{P} = P$. Since $z\widehat{J} = zP$, we obtain $zI \subseteq z\widehat{I} \subseteq P$, and hence $z \in I^{-1} = zJ$, a contradiction.

This shows that $\mathcal{M}(P)$ is a group and that $\beta)$ holds. For $I_1 \supseteq I_2$ in $\mathcal{M}(P)$ we define $I_1 \leq I_2$ and $\mathcal{M}(P)$ then is a linearly ordered group with P as identity. Elements in $\mathcal{H}(P)$ have the form $I = zP = Pz$ for some z in G with $z\widehat{P} = zP$ and $(zP)^{-1} = z^{-1}P = Pz^{-1}$; see f) in Section 1.1 and $\gamma)$ follows. This proves the theorem. \square

Corollary 1.8. *Let P be a cone of rank one in a group G . Then $\mathcal{M}(P)$ and $\mathcal{H}(P)$ are Archimedean groups.*

Proof. Let $B \subset P$ be a divisorial ideal. If $B \subset J(P) = J$ or $B = J \neq J^2$, then $\bigcap B^n = \emptyset$ by Property d) in Section 1.1. If $J = J^2$, we have $\widehat{J} = P$ and hence $\bigcap B^n = \emptyset$ in all cases, and $B^{n+1} \subset B^n$. Then $\widehat{B^{n+1}} \subseteq B^n$, since there are no further right ideals between B^{n+1} and $\widehat{B^{n+1}}$. This implies $\bigcap \widehat{B^n} = \emptyset$, and it follows that $\mathcal{M}(P)$ and $\mathcal{H}(P)$ are Archimedean; see also Property f) in Section 1.1. \square

Related results can be found in [12] and [2].

1.3. The classification of rank one cones. The groups $\mathcal{M}(P)$ and $\mathcal{H}(P)$ will be used to classify rank one cones P in groups G based on the lattice of their ideals. In the following theorem and proof we will write J instead of $J(P)$.

Theorem 1.9. *Let P be a cone of rank one in a group G . Then exactly one of the following possibilities occurs:*

A) : *The cone P is Archimedean, i.e., $aP = Pa$ for all a in P . We distinguish two possibilities in this case:*

A₁): $\mathcal{M}(P) = \mathcal{H}(P) \cong (\mathbb{Z}, +, \leq)$, which is exactly the case when $J^2 \neq J$. Then every P -ideal is a power of J and the cone is called discrete.

A₂): $\mathcal{M}(P) \cong (\mathbb{R}, +, \leq)$ and $\mathcal{H}(P)$ is a dense subgroup of $\mathcal{M}(P)$.

B) : *The cone P is nearly simple; i.e., J is the only proper ideal in P . In this case $\mathcal{M}(P) = \mathcal{H}(P) = \{P\}$.*

C) : *The cone P is exceptional; i.e., there exists a prime ideal Q in P that is not completely prime. Then:*

i) : *There are no further ideals between J and Q .*

ii) : *The ideal Q is divisorial and $\mathcal{M}(P) = \text{gr}\{Q\}$ is an infinite cyclic group.*

iii) : $\bigcap Q^n = \emptyset$.

iv) : *There exists an integer $k \geq 0$ such that $\mathcal{H}(P) = \text{gr}\{\widehat{Q^k}\}$. The cone P is said to be of type (C_k) in this case.*

If P is of type (C_0) , then

$$\dots \supset (Q^n)^{-1} \supset \dots \supset Q^{-1} \supset P \supset J \supset Q \supset Q^2 \supset Q^3 \supset \dots$$

is the chain of P -ideals.

If P is of type (C_1) , then

$$\begin{aligned} \dots \supset Q^{-2} &= z^{-2}P \supset z^{-2}J \supset Q^{-1} = z^{-1}P \supset z^{-1}J \supset P \supset J \supset zP \\ &= Q \supset zJ \supset z^2P = Q^2 \supset z^2J \supset \dots \end{aligned}$$

is the chain of P -ideals.

If P is of type (C_k) , $k \geq 2$, then

$$\begin{aligned} \dots \supset (Q^{k+1})^{-1} \supset z^{-1}P \supset z^{-1}J \supset (Q^{k-1})^{-1} \supset \dots \\ \supset Q^{-1} \supset P \supset J \supset Q \supset Q^2 \supset \dots \supset Q^{k-1} \supset zP \supset zJ \\ = Q^k \supset Q^{k+1} \supset \dots \supset Q^{2k-1} \supset z^2P \supset z^2J \\ = Q^{2k} \supset Q^{2k+1} \supset \dots \end{aligned}$$

is the chain of all P -ideals.

Proof. If J is the only proper ideal of P , then P is of type B .

Otherwise, let $Q = \bigcup I$ be the union of ideals of P properly contained in J . If $J^2 = J$ and $J \supset Q$, then P is exceptional: for ideals $I_1 \supset Q$ and $I_2 \supset Q$ in P we have $I_1 \cdot I_2 \supseteq J^2 = J \supset Q$ and Q is a prime ideal of P , not completely prime and no further ideal exists between J and Q . The divisorial closure \widehat{Q} of Q is an ideal that cannot be equal to J , since J would then be right principal. Hence, $\widehat{Q} = Q$ is the smallest positive element in the linearly ordered Archimedean group $\mathcal{M}(P)$, and $\mathcal{M}(P) = \text{gr}\{Q\}$ is an infinite cyclic group. The subgroup $\mathcal{H}(P)$ has therefore the form $\mathcal{H}(P) = \text{gr}\{\widehat{Q}^k\}$ for some $k \geq 0$; we say that P is of type (C_k) .

We can now describe the P -ideals in each case (C_k) if we recall (see Property d) in Section 1.2) that an ideal I is either divisorial or of the form $cJ = Jc$ with $\widehat{I} = cP = Pc$, some $c \in G$ and $J = J^2$. It will also follow from the rest of the proof that if P is exceptional, then $J = J^2$ and $J \supset Q = \bigcup I$, where the ideals I are properly contained in J , the prime ideal that is not completely prime.

In the case (C_0) there are no principal ideals $\neq P$ and the group $\mathcal{M}(P) = \text{gr}\{Q\}$ contains all P -ideals $\neq J$. In the case (C_1) the ideal $Q = zP = Pz$ is principal and $\mathcal{M}(P) = \mathcal{H}(P)$. In the case (C_k) , $k \geq 2$, the ideal \widehat{Q}^k is principal. However, Q^k itself cannot be principal, since otherwise $Q^k = zP$ implies $Q^k J \neq Q^k$; hence, $QJ \neq Q$ and Q is principal (see Property g) in Section 1.1). Hence $\widehat{Q}^k = zP = Pz \supset Q^k = zJ = Jz$ for an element z in P .

It remains to consider the case where either $J \neq J^2$, or $J = J^2$ and $J = Q = \bigcup I$ for ideals I properly contained in J . In this case we will prove that $aP = Pa$ for all a in P . If for some a in P the right ideal aP is not a left ideal, then an element c exists in P with $caP \supset aP$ and $caj = a$ follows for an element j in J . By assumption there exists an ideal $I \subseteq J$ with $j \in I$ and $\bigcap I^n = \emptyset$; we obtain the contradiction $a = caj = c^n a j^n \in \bigcap I^n$. We have $Pa \subseteq aP$, $P \subseteq aPa^{-1}$ and $P = aPa^{-1}$ since P is of rank one. Therefore, $Pa = aP$ for all a in P and P is invariant.

If $J \neq J^2$, then $J = aP = Pa$, for some a in P , is the smallest positive element in the Archimedean group $\mathcal{M}(P)$. Hence, $\mathcal{M}(P) = \mathcal{H}(P) = \text{gr}\{J\}$ is the group of all P -ideals.

If $J = J^2$ and $J = Q$, then $\mathcal{H}(P)$ is isomorphic to a dense subgroup of $(\mathbb{R}, +, \leq)$ and $\mathcal{M}(P)$ is isomorphic to $(\mathbb{R}, +, \leq)$. \square

If R is a chain order of rank one in a skew field D , then $R^* = R \setminus \{0\}$ is a cone in the group D^* . We say that R has type (A) , (A_1) , (A_2) , (B) , (C) , or (C_k) if and only if the cone R^* is of the same type.

The next result follows from Propositions 1.3 and 1.5 and Theorem 1.9.

Corollary 1.10. *Let P be a cone associated with the rank one chain domain R . Then P and R have the same type.*

2. THE UNIVERSAL COVERING GROUP \mathbb{G} OF $\mathrm{SL}(2, \mathbb{R})$

2.1. The group $\mathrm{SL}(2, \mathbb{R})$. By $\mathrm{SL}(2, \mathbb{R})$ we denote, as usual, the group of 2×2 matrices with real entries and determinant equal to 1. Then

$$\mathbb{U} = \left\{ u = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mid a, b \in \mathbb{R}, a > 0 \right\}$$

and

$$\mathbb{S} = \left\{ r(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

are two particular subgroups of \mathbb{G} . Every element $s \in \mathrm{SL}(2, \mathbb{R})$ can be written in a unique way as

$$s = r(t)u \quad \text{for } r(t) \in \mathbb{S} \quad \text{with } 0 \leq t < 2\pi \quad \text{and } u \in \mathbb{U}.$$

To prove this claim, let $\{e_1, e_2\}$ be the standard basis of \mathbb{R}^2 , the Euclidean plane, and let the elements of $\mathrm{SL}(2, \mathbb{R})$ be the representations of linear transformations of \mathbb{R}^2 with respect to the basis $\{e_1, e_2\}$. For every nonzero vector $\mathbf{a} \in \mathbb{R}^2$ there exists a unique element $t \in [0, 2\pi)$ with $\mathbf{a}/\|\mathbf{a}\| = e_1 \cos t + e_2 \sin t$; we write $\arg \mathbf{a} = t$ in this case.

Let $t = \arg s(e_1)$ for the given element $s \in \mathrm{SL}(2, \mathbb{R})$ and $r(-t)s = u \in \mathbb{U}$ for some element u , since $r(-t)s(e_1) = ae_1$ for $a > 0$. Hence, $s = r(t)u$ and this representation is unique, since $\mathbb{U} \cap \mathbb{S} = \{I\}$, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the identity of $\mathrm{SL}(2, \mathbb{R})$.

2.2. The group \mathbb{G} . We are going to construct the universal covering group \mathbb{G} of the group $\mathrm{SL}(2, \mathbb{R})$ in this section. We do this first for the subgroup \mathbb{S} by fixing a symbol, say x , and by rewriting the additive group of the real numbers in multiplicative form:

$$R = \{x^t \mid t \in \mathbb{R}\}; \quad x^{t_1} \cdot x^{t_2} = x^{t_1+t_2}; \quad x^{t_1} \leq x^{t_2} \Leftrightarrow t_1 \leq t_2.$$

Then R is a linearly ordered group isomorphic to $(\mathbb{R}, +, \leq)$. The mapping τ from R to \mathbb{S} with $\tau(x^t) = r(t)$ is a group epimorphism with the cyclic subgroup $\mathrm{gr}\{x^{2\pi}\}$ as its kernel; τ is a cover of the Lie group \mathbb{S} . Next we define the covering group \mathbb{G} of $\mathrm{SL}(2, \mathbb{R})$ as the set $\mathbb{G} = \{x^t u \mid x^t \in R, u \in \mathbb{U}\}$, the Cartesian product $R \times \mathbb{U}$, together with the following operation: If $x^{t_1} u_1, x^{t_2} u_2$ are two elements in \mathbb{G} and $t_2 = 2\pi k + \varphi$ for $k \in \mathbb{Z}$ and $\varphi \in [0, 2\pi)$, then $u_1 r(\varphi) u_2 = r(\psi) u$ in $\mathrm{SL}(2, \mathbb{R})$ for $u \in \mathbb{U}$, $\psi \in [0, 2\pi)$, and the product in \mathbb{G} is defined as $x^{t_1} u_1 \cdot x^{t_2} u_2 = x^{t_1+2\pi k+\psi} u$.

The mapping τ from above can be extended to a mapping from \mathbb{G} to $\mathrm{SL}(2, \mathbb{R})$ by defining

$$\tau(x^t u) = r(t)u.$$

We want to prove that \mathbb{G} is a group and that τ is an epimorphism from \mathbb{G} onto $\mathrm{SL}(2, \mathbb{R})$.

Lemma 2.1. *The mapping τ is onto $\mathrm{SL}(2, \mathbb{R})$, and if $a \cdot b = c$ for elements a, b, c in \mathbb{G} , then $\tau(c) = \tau(a)\tau(b)$.*

Proof. The element $x^t u$ in \mathbb{G} satisfies $\tau(x^t u) = r(t)u$ for the arbitrary element $r(t)u$ in $\text{SL}(2, \mathbb{R})$; τ is onto. If $a = x^{t_1} u_1$, $b = x^{t_2} u_2$ in \mathbb{G} , $t_2 = 2\pi k + \varphi$, $k \in \mathbb{Z}$, $\varphi \in [0, 2\pi)$ and if $u_1 r(\varphi) u_2 = r(\psi)u$, $\psi \in [0, 2\pi)$, $u_i, u \in \mathbb{U}$, then $c = x^{t_1 + 2\pi k + \psi} u$ and

$$\begin{aligned}\tau(c) &= r(t_1 + 2\pi k + \psi)u = r(t_1)r(\psi)u = r(t_1)u_1 r(\varphi)u_2 \\ &= \tau(a)r(2\pi k + \varphi)u_2 = \tau(a)\tau(b),\end{aligned}$$

which proves the lemma. \square

Several special cases of the associative law for the operation defined for \mathbb{G} are proved in the next few steps. We can consider R as well as \mathbb{U} as subgroups of \mathbb{G} and the equations

$$(+)\quad x^t \cdot u = x^t u, \quad x^{t_1} \cdot x^t u = x^{t_1+t} u, \quad \text{and} \quad x^t u \cdot u' = x^t u u'$$

follow. We conclude also that $x^t \cdot a = x^t \cdot b$ implies $a = b$ for elements $a, b \in \mathbb{G}$.

Lemma 2.2. *For any element $g = x^t u \in \mathbb{G}$ and any $m \in \mathbb{Z}$ the product $g \cdot x^{\pi m}$ is equal to $x^{t+\pi m} u$.*

Proof. We have $\pi m = 2\pi k + \varphi$ with $k \in \mathbb{Z}$, and $\varphi = 0$ if m is even, and $\varphi = \pi$ if m is odd. In both cases $ur(\varphi) = r(\varphi)u$ follows, which proves the statement of the lemma. \square

Lemma 2.3. *For any $a, b \in \mathbb{G}$ and any integer $m \in \mathbb{Z}$ the following equalities hold:*

$$x^{\pi m} \cdot (a \cdot b) = (x^{\pi m} \cdot a) \cdot b = a \cdot (x^{\pi m} \cdot b).$$

Proof. Because of (+) the first equation follows, and we can assume that $a = u \in \mathbb{U}$ and $b = x^t \in R$ in the second equation.

It remains to prove the following equality:

$$(x^{\pi m} \cdot u) \cdot x^t = u \cdot (x^{\pi m} \cdot x^t)$$

where $t = 2\pi k + \varphi$, $k \in \mathbb{Z}$, $\varphi \in [0, 2\pi)$ and $ur(\varphi) = r(\varphi)u$ for $\varphi \in [0, 2\pi)$, $u' \in \mathbb{U}$.

Then $(x^{\pi m} \cdot u) \cdot x^t = x^{\pi m + 2\pi k + \psi} u'$. We distinguish three cases in order to compute the right-hand side of the above equation.

In the first case, $m = 2k'$ is even and the equality follows immediately.

In the second case, $\pi m = 2\pi k' + \pi$ for some $k' \in \mathbb{Z}$ and $\varphi < \pi$. Then

$$u \cdot x^{2\pi(k+k')+\varphi+\pi} = x^{2\pi(k+k')} u \cdot x^{\pi+\varphi} = x^{2\pi(k+k')+\pi+\psi} u' = x^{\pi m + 2\pi k + \psi} u'$$

since $ur(\pi + \varphi) = ur(\pi)r(\varphi) = r(\pi)r(\varphi)u' = r(\pi + \psi)u'$ in $\text{SL}(2, \mathbb{R})$; the equation is proved in this case.

In the final case, $\pi m = 2\pi k' + \pi$ for $k' \in \mathbb{Z}$ and $\varphi \geq \pi$. The right-hand side of the above equation is then equal to

$$u \cdot x^{2\pi(k+k'+1)+\varphi-\pi} = x^{2\pi(k+k'+1)-\pi+\psi} u' = x^{\pi m + 2\pi k + \psi} u',$$

which proves the lemma. \square

Lemma 2.4. *Let $u \in \mathbb{U}$ and $t \in (\pi m, \pi(m+1))$ for some $m \in \mathbb{Z}$. Then $u \cdot x^t = x^{t'} u'$ for $u' \in \mathbb{U}$ and $t' \in (\pi m, \pi(m+1))$.*

Proof. Let $t = 2\pi k + \varphi$ for $k \in \mathbb{Z}$, $\varphi \in (0, 2\pi)$. If $m = 2k$ is even, then $\varphi \in (0, \pi)$; hence $\sin \varphi > 0$. It follows that for any $u = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \in \mathbb{U}$; the argument ψ of $ur(\varphi)(e_1) = r(\varphi)u'(e_1)$ is also in $(0, \pi)$ since $\psi = \arg \left[\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \right]$ and $a^{-1} \sin \varphi > 0$. Hence $t' = 2\pi k + \psi \in (\pi m, \pi(m+1))$ as stated in the lemma.

If $m = 1 + 2k$ is odd, then $t = 2\pi k + \varphi$ and $\varphi \in (\pi, 2\pi)$. Then $\sin \varphi < 0$ and $\sin \psi$ with $ur(\varphi) = r(\psi)u'$ is also negative with the above argument; hence, $\psi \in (\pi, 2\pi)$ and $t' = 2\pi k + \psi \in (\pi m, \pi(m + 1))$. \square

Theorem 2.5. a): \mathbb{G} is a group;

b): The mapping τ is a homomorphism from \mathbb{G} onto $\text{SL}(2, \mathbb{R})$;

c): The center of \mathbb{G} is the infinite cyclic group generated by x^π .

Proof. To show that the operation defined for \mathbb{G} is associative we consider three elements $x^{t_i}u_i \in G$, $i = 1, 2, 3$ with $t_i \in \mathbb{R}$ and $u_i \in \mathbb{U}$ and the equation

$$(*) \quad g_1 = (x^{t_1}u_1 \cdot x^{t_2}u_2) \cdot x^{t_3}u_3 = x^{t_1}u_1 \cdot (x^{t_2}u_2 \cdot x^{t_3}u_3) = g_2.$$

By Lemmas 2.2 and 2.3 this equation holds if and only if the following equation is true:

$$(x^{t_1+\pi k}u_1 \cdot x^{t_2+\pi m}u_2) \cdot x^{t_3+\pi n}u_3 = x^{t_1+\pi k}u_1 \cdot (x^{t_2+\pi m}u_2 \cdot x^{t_3+\pi n}u_3)$$

for integers k, m and n .

It follows that it is sufficient to prove $(*)$ only in the case where $t_1, t_2, t_3 \in [0, \pi)$.

For $g_1 = x^t u'$ and $g_2 = x^{t'} u''$ with $t, t' \in \mathbb{R}$, $u', u'' \in \mathbb{U}$ we apply Lemma 2.1 and obtain

$$r(t)u' = (r(t_1)u_1 \cdot r(t_2)u_2) \cdot r(t_3)u_3$$

and

$$r(t')u'' = r(t_1)u_1 \cdot (r(t_2)u_2 \cdot r(t_3)u_3)$$

in $\text{SL}(2, \mathbb{R})$ where the operation is associative, and therefore $r(t)u' = r(t')u''$ follows. This implies $u' = u''$ and $t - t' = 2\pi k$ for some $k \in \mathbb{Z}$. It remains to show that $k = 0$.

We apply Lemma 2.4 and obtain $u_1 x^{t_2} = x^{t'_2} u'_1$ for $u'_1 \in \mathbb{U}$, $t'_2 \in [0, \pi)$; $u'_1 u_2 \cdot x^{t_3} = x^{\tilde{t}_3} \tilde{u}$ for $\tilde{u} \in \mathbb{U}$, $\tilde{t}_3 \in [0, \pi)$; $u_2 x^{t_3} = x^{t'_3} u'_2$ for $u'_2 \in \mathbb{U}$, $t'_3 \in [0, \pi)$; and $u_1 x^{t_2+t'_3} = x^{t_{2,3}} u''_1$ for $u''_1 \in \mathbb{U}$ and $t_{2,3} \in [0, \pi)$.

Therefore:

$$\begin{aligned} g_1 &= (x^{t_1}u_1 \cdot x^{t_2}u_2) \cdot x^{t_3}u_3 = (x^{t_1+t'_2}u'_1 u_2) \cdot x^{t_3}u_3 \\ &= x^{t_1+t'_2+\tilde{t}_3} \tilde{u} u_3, \end{aligned}$$

and

$$\begin{aligned} g_2 &= x^{t_1}u_1 \cdot (x^{t_2}u_2 \cdot x^{t_3}u_3) = x^{t_1}u_1 \cdot x^{t_2+t'_3}u'_2 u_3 \\ &= x^{t_1+t_{2,3}} u''_1 u'_2 u_3. \end{aligned}$$

Hence, $t = t_1 + t'_2 + \tilde{t}_3$ and $t' = t_1 + t_{2,3}$ and therefore

$$t - t' = t'_2 + \tilde{t}_3 - t_{2,3} = 2\pi k.$$

However, $t'_2 + \tilde{t}_3$ and $t_{2,3}$ both belong to $[0, 2\pi)$ and $k = 0$ and the associative law follows for the operation defined for \mathbb{G} .

Since \mathbb{G} has $e = x^0 E$, for $E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, as the identity and $x^t u$ has $u^{-1} x^{-t}$ as its inverse, \mathbb{G} is indeed a group; this proves a).

The statement b) was proved in Lemma 2.1. It follows from Lemma 2.3 that $\text{gr}\{x^\pi\}$ is contained in the center $Z(\mathbb{G})$ of \mathbb{G} . Conversely, if $x^t u \in Z(\mathbb{G})$ for $t \in \mathbb{R}$ and $u \in \mathbb{U}$, then an application of Lemma 2.1 shows that $r(t)u$ is in $Z(\text{SL}(2, \mathbb{R}))$.

Hence $r(t)u = \pm r(0)$, $u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $t = \pi k$ for some $k \in \mathbb{Z}$ follows. Therefore $x^t u \in \text{gr}\{x^\pi\}$, which proves c) and the theorem. \square

See also [1] for the fact that \mathbb{G} is right orderable, but not locally indicable.

2.3. The representation of the group \mathbb{G} . To each element $g = x^t u \in \mathbb{G}$ we can assign the projection $v(g) = v(x^t u) = x^t \in R$. The mapping $V : \mathbb{G} \rightarrow \text{Aut}(R, \leq)$ is defined as $V_g(x^t) = v(gx^t)$ for $g \in \mathbb{G}$, $x^t \in R$. That V_g is indeed an automorphism of (R, \leq) follows from the next result.

Lemma 2.6. *For $g \in \mathbb{G}$ let V_g be defined as above. Then:*

- a): $V_{g_1 g_2} = V_{g_1} \circ V_{g_2}$ for $g_1, g_2 \in \mathbb{G}$.
- b): V_g is the identity mapping if and only if g is the identity element in \mathbb{G} .
- c): The stabilizer $\text{st}(x^t) = \{g \in \mathbb{G} \mid V_g(x^t) = x^t\}$ is equal to $x^t \mathbb{U} x^{-t} \cong \mathbb{U}$, which is an Ore group.
- d): V_g is an automorphism of (R, \leq) for every $g \in \mathbb{G}$.

Proof. To prove a) we compute $v(g_1 g_2 x^t)$ and $v(g_1 v(g_2 x^t))$. Let $g_1 = x^{t_1} u_1$, $g_2 = x^{t_2} u_2$ for $u_i \in \mathbb{U}$. Then $g_1 g_2 x^t = x^{t_1} u_1 x^{t_2} u_2 x^t = x^{t_1} u_1 x^{t_2} x^{t'} u'$ for some $u' \in \mathbb{U}$, $t' \in \mathbb{R}$ with $u_2 x^t = x^{t'} u'$. Further, $x^{t_1} u_1 x^{t_2+t'} u' = x^{t_1+\tilde{t}} \tilde{u} u'$ for $u_1 x^{t_2+t'} = x^{\tilde{t}} \tilde{u}$ for $\tilde{u} \in \mathbb{U}$, $\tilde{t} \in \mathbb{R}$. It follows that $v(g_1 g_2 x^t) = x^{t_1+\tilde{t}}$ and that $v(g_1 v(g_2 x^t)) = v(x^{t_1} u_1 x^{t_2+t'}) = x^{t_1+\tilde{t}}$; this proves a).

To prove b), assume $g = x^{t_1} u$ and $V_g(x^t) = x^t$ for all $t \in \mathbb{R}$. For $t = 0$ it follows that $t_1 = 0$. We consider $t = \frac{\pi}{2}$ and assume that $u = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$. Then $V_u(x^{\frac{\pi}{2}}) = x^{\frac{\pi}{2}}$ implies that

$$\arg \left[\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right] = \arg \begin{pmatrix} b \\ a^{-1} \end{pmatrix} = \frac{\pi}{2}.$$

Hence, $b = 0$ and $u = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$. Finally, for $t = \frac{\pi}{4}$ we must have

$$\arg \left[\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right] = \frac{\pi}{4}$$

and $a = a^{-1} = 1$ follows; hence, $g = e$, the identity in \mathbb{G} , and b) follows.

To prove c) we observe that $\text{st}(x^0) = \{x^{t_1} u_1 \in \mathbb{G} \mid V_g(x^0) = x^{t_1} = x^0\}$ equals \mathbb{U} . Hence, $V_g(x^t) = x^t \mathbb{U} x^{-t} \cong \mathbb{U}$. These stabilizers are Ore groups in the sense that the group ring $T\mathbb{U}$ over a skew field T is an Ore domain. This is true since \mathbb{U} is the semidirect product of the following two torsion free abelian groups:

$$A = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \mid 0 < a \in \mathbb{R} \right\} \quad \text{and} \quad B = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid b \in \mathbb{R} \right\}.$$

This proves c).

Finally, we want to prove d). Since V_{x^t} is an automorphism of (R, \leq) , it follows from a) that it is enough to show that V_u is an automorphism of (R, \leq) for any $u \in \mathbb{U}$. We show first that $x^{t_2} > x^{t_1}$ implies $V_u(x^{t_2}) > V_u(x^{t_1})$ which then implies that V_u is one-to-one and order-preserving. By Lemma 2.4 and Theorem 2.5(c) we can assume $t_1, t_2 \in [0, \pi)$. It then follows that $t_2 - t_1 \in (-\pi, \pi)$, and in addition $t_2 - t_1 > 0$ if and only if

$$\sin(t_2 - t_1) = \text{Det} \begin{pmatrix} \cos t_1 & \cos t_2 \\ \sin t_1 & \sin t_2 \end{pmatrix} > 0.$$

We have $ur(t_i) = r(d_i)u_i$ for $d_i \in [0, \pi)$ and $d_i = \arg(u(\begin{pmatrix} \cos t_i \\ \sin t_i \end{pmatrix})) \in [0, \pi)$. Then $\text{Det}(u(\begin{pmatrix} \cos t_1 & \cos t_2 \\ \sin t_1 & \sin t_2 \end{pmatrix})) > 0$, since $\text{Det}(u) > 0$, and, as in the previous argument,

$d_2 > d_1$ follows. This shows that $x^{d_2} = V_u(x^{t_2}) > x^{d_1} = V_u(x^{t_1})$ for $t_2 > t_1$ and that V_u is order-preserving and one-to-one.

It remains to show that V_u is onto, and by Lemma 2.4 and Theorem 2.5(c) it is enough to show that V_u maps the interval $[x^0, x^\pi]$ onto the interval $[x^0, x^\pi]$. This, however, follows from the fact that $V_u(x^0) = x^0$, $V_u(x^\pi) = x^\pi$ and that V_u is continuous. \square

We will prove next a technical result which will be used several times.

Lemma 2.7. *Let $g = x^t u \in \mathbb{G}$ with $t = \pi k + t_0$ and $x^{t_1} \in R$ with $t_1 = \pi m + t_{10}$ for $k, m \in \mathbb{Z}$ and $t_0, t_{10} \in [0, \pi)$. Assume that $(\frac{a}{b}) \in \mathbb{R}^2$ with $\arg(\frac{a}{b}) = t_{10}$.*

Then $V_g(x^{t_1}) = x^{\pi(k+m)+t'}$ for $t' = \arg(r(t_0)u(\frac{a}{b}))$.

Proof. By definition we have that $V_g(x^{t_1}) = v(gx^{t_1})$. Further, $gx^{t_1} = x^t u x^{t_1} = x^{\pi(k+m)} x^{t_0} u x^{t_{10}}$ since x^π is in the center of \mathbb{G} by Theorem 2.5(c).

By Lemma 2.4 we have $u x^{t_{10}} = x^{\tilde{t}} \tilde{u}$ with $\tilde{u} \in \mathbb{U}$ and $\tilde{t} = \arg(ur(t_{10})(\frac{1}{0})) \in [0, \pi)$. Hence, $x^{t_0} u x^{t_{10}} = x^{t_0+\tilde{t}} \tilde{u}$. On the other hand, $t' = \arg(r(t_0)u(\frac{a}{b})) = t_0 + \tilde{t}$, since both $t_0, \tilde{t} \in [0, \pi)$. It follows that $gx^{t_1} = x^{\pi(k+m)+t_0+\tilde{t}} \tilde{u}$ and $V_g(x^{t_1}) = x^{\pi(k+m)+t'}$. \square

3. EXCEPTIONAL CONES IN THE UNIVERSAL COVERING GROUP \mathbb{G}

In this section we construct exceptional cones of type (C_k) for every k in the universal covering group \mathbb{G} of $\text{SL}(2, \mathbb{R})$.

We define first two particular elements w_1, w_2 in \mathbb{G} which will play an important role in this construction. The element $w_1 = (\frac{1}{0} \frac{2}{1}) \in \mathbb{U} \subset \mathbb{G}$ and $\tau(w_1) = w_1$ follows. Next we consider the element $(\frac{1}{2} \frac{0}{1}) \in \text{SL}(2, \mathbb{R})$ and $\alpha = \arg[(\frac{1}{2} \frac{0}{1})(\frac{1}{0})] = \arctan 2 \in (0, \pi)$ and define w_2 as $x^\alpha u$ where $u = r(-\alpha)(\frac{1}{2} \frac{0}{1}) \in \mathbb{U}$; hence, $\tau(w_2) = (\frac{1}{2} \frac{0}{1})$.

Lemma 3.1. *Let b be an element in $[0, \pi)$. Then $\lim_{n \rightarrow \infty} V_{w_1^n}(x^b) = x^0$.*

Proof. We consider the real number b_n with $x^{b_n} = V_{w_1^n}(x^b)$. Since $w_1^n = (\frac{1}{0} \frac{2n}{1})$ and $\tau(w_1^n) = w_1^n$, we can apply Lemma 2.7 and obtain

$$b_n = \arg[(\frac{1}{0} \frac{2n}{1})(\frac{\cos b}{\sin b})] = \arg(\frac{\cos b + 2n \sin b}{\sin b}).$$

If $b = 0$, then $b_n = 0$ for all $n \geq 0$ and the result follows. If $b \in (0, \pi)$, then $\sin b > 0$ and $\lim_{n \rightarrow \infty} (\cos b + 2n \sin b) = \infty$; the statement of the lemma follows. \square

We are now ready to define one of the main objects of this paper:

$$\mathbb{P} = \{g \in \mathbb{G} \mid V_g(x^0) \geq x^0\}.$$

The next result shows that this is an exceptional cone of type (C_1) in \mathbb{G} .

Theorem 3.2. a): *The set $\mathbb{P} = \{g \in \mathbb{G} \mid V_g(x^0) \geq x^0\}$ is a cone in \mathbb{G} with $U(\mathbb{P}) = \mathbb{U}$.*

- b):** *Any right \mathbb{P} -ideal is either a principal right ideal $x^t \mathbb{P}$ or of the form $x^t J(\mathbb{P})$ for some $t \in \mathbb{R}$.*
- c):** *Any \mathbb{P} -ideal has the form $x^{\pi m} \mathbb{P}$ or $x^{\pi m} J(\mathbb{P})$ for some m in \mathbb{Z} .*
- d):** *The cone \mathbb{P} is exceptional of rank one with $Q = x^\pi \mathbb{P}$ the prime ideal that is not completely prime; \mathbb{P} is exceptional of type (C_1) .*

Proof. a) If g and h are elements in \mathbb{P} , then $V_{gh}(x^0) = V_g(V_h(x^0)) \geq V_g(x^0) \geq x^0$ by Lemma 2.6, a) and d), and $gh \in \mathbb{P}$ follows.

If g is not in \mathbb{P} , then $V_g(x^0) < x^0$; hence, $x^0 < V_{g^{-1}}(x^0)$ again by Lemma 2.6, and $g^{-1} \in \mathbb{P}$ follows and \mathbb{P} is a cone of \mathbb{G} . It also follows from the above arguments that $g, g^{-1} \in \mathbb{P}$ implies $V_g(x^0) = x^0$ and $g \in \mathbb{U}$. Conversely, $\mathbb{U} \subset \mathbb{P}$ and $U(\mathbb{P}) = \mathbb{U}$ follows. Hence, $J(\mathbb{P}) = \{g \in \mathbb{G} \mid V_g(x^0) > x^0\}$.

b) Let I be any right \mathbb{P} -ideal in \mathbb{G} . Then it follows that $x^\alpha = \inf\{V_g(x^0) \mid g \in I\}$ exists since $I \subseteq c\mathbb{P}$ for some $c \in \mathbb{G}$. We will show that $\widehat{I} = x^\alpha\mathbb{P}$ for the divisorial closure \widehat{I} of I , see Definition 1.4. By definition we have $x^\alpha\mathbb{P} \supseteq g\mathbb{P} = x^\beta\mathbb{P}$ for all $g \in I$ since $\alpha \leq \beta$; hence $x^\alpha\mathbb{P} \supseteq I$. Conversely, if $h \in \mathbb{G}$ with $h\mathbb{P} = x^\gamma\mathbb{P} \supseteq I$, then $\gamma \leq V_g(x^0)$ for all $g \in I$ and $\gamma \leq \alpha$ follows; hence $\widehat{I} = x^\alpha\mathbb{P}$. It follows that either $I = x^\alpha\mathbb{P} = \widehat{I}$ or that $\widehat{I} = x^\alpha\mathbb{P}$ and $I = x^\alpha J(\mathbb{P})$; see Property d) in Section 1.2.

c) Assume that $x^t\mathbb{P}$ is a \mathbb{P} -ideal. For $t = \pi m + t_0$, $m \in \mathbb{Z}$ and $t_0 \in [0, \pi)$ it follows that $x^{t_0}\mathbb{P}$ is also a \mathbb{P} -ideal since x^π is central in \mathbb{G} . If $t_0 > 0$, it follows from Lemma 3.1 that there exists a power w_1^n of w_1 in \mathbb{U} with $w_1^n x^{t_0}\mathbb{P} \supset x^{t_0}\mathbb{P}$, a contradiction that shows that $x^t\mathbb{P} = x^{\pi m}\mathbb{P}$. If $I = x^t J(\mathbb{P})$ is a \mathbb{P} -ideal, then $\widehat{I} = x^t\mathbb{P}$ is a \mathbb{P} -ideal by Property d) in Section 1.2, and $t = \pi m$ by the above argument.

d) We have $\mathbb{P} \supset J(\mathbb{P}) \supset x^\pi\mathbb{P} = Q$ and Q is not a completely prime ideal of \mathbb{P} , since $x^{\pi/2} \cdot x^{\pi/2} \in Q$, but $x^{\pi/2} \notin Q$. However, Q is a prime ideal, since any ideals A and B of \mathbb{P} that contain Q properly, also contain $J(\mathbb{P})$; hence, $AB \supseteq J(\mathbb{P})J(\mathbb{P}) = J(\mathbb{P}) \supset Q$, and it follows that Q is a prime ideal that is not completely prime. There are no further ideals between $J(\mathbb{P})$ and Q , and $\bigcap Q^n = \emptyset$. It follows that \mathbb{P} is an exceptional cone of type (C_1) ; see Theorem 1.9. \square

We denote by F the subgroup $\text{gr}\{w_1, w_2\}$ of \mathbb{G} generated by w_1 and w_2 . This subgroup is mapped by τ onto the subgroup $\text{gr}\left\{\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\right\}$ of $\text{SL}(2, \mathbb{R})$ generated by $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Since this subgroup of $\text{SL}(2, \mathbb{R})$ is free (see [15], 14.2.1), the group F is free of rank 2.

Lemma 3.2. *Let h_1 be the element $w_2 w_1^{-1} w_2$ in F . Then $V_{h_1}(x^0) = x^{\arg \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \pi} \in (x^\pi, x^{\pi + \frac{\pi}{2}})$.*

Proof. We have $w_2 x^0 = x^\alpha u$ for $\alpha = \arg \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in (0, \pi)$ with

$$u = r(-\alpha) \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \in \mathbb{U}.$$

It follows from Lemma 2.7 that

$$V_{w_1^{-1}}(x^\alpha) = x^{t'} = x^{\arg \begin{pmatrix} -3 \\ 2 \end{pmatrix}}, \text{ since } w_1^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \in \mathbb{U}$$

and

$$t' = \arg \left[\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right] = \arg \begin{pmatrix} -3 \\ 2 \end{pmatrix}.$$

By a further application of Lemma 2.7 we obtain

$$V_{h_1}(x^0) = V_{w_2} \left(x^{\arg \begin{pmatrix} -3 \\ 2 \end{pmatrix}} \right) = x^{t''}$$

with

$$t'' = \arg [\tau(w_2) \begin{pmatrix} -3 \\ 2 \end{pmatrix}] = \arg \left[\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ 2 \end{pmatrix} \right] = \arg \begin{pmatrix} -3 \\ -4 \end{pmatrix}.$$

Hence, $V_{h_1}(x^0) = x^{\arg \begin{pmatrix} -3 \\ -4 \end{pmatrix}} = x^{\pi + \arg \begin{pmatrix} 3 \\ 4 \end{pmatrix}}$, which proves the lemma. \square

In order to construct further cones we consider a subgroup H of \mathbb{G} that contains F and define

$$P_H = H \cap \mathbb{P}.$$

It follows immediately that P_H is closed under multiplication. If $g \in H \setminus P_H$, then $g \notin \mathbb{P}$ and $g^{-1} \in H \cap \mathbb{P} = P_H$ follows; P_H is a cone of H .

Lemma 3.3. *Let $t \in \mathbb{R}$. Then $P_H x^t \mathbb{P} = \mathbb{P} x^t \mathbb{P}$.*

Proof. It is enough to prove this for $t \in [0, \pi)$, since $t = k\pi + t_0$, $t_0 \in [0, \pi)$ in the general case with x^π in the center of G .

If $t = 0$, then $P_H x^t \mathbb{P} = \mathbb{P} = \mathbb{P} x^t \mathbb{P}$. If $t \in (0, \pi)$, then for any $j \in J(\mathbb{P})$ there exists an n with $w_1^n x^t \mathbb{P} \supseteq j\mathbb{P}$ by Lemma 3.1. Hence, $\mathbb{P} x^t \mathbb{P} = J(\mathbb{P}) = \bigcup w_1^n x^t \mathbb{P} \subseteq P_H x^t \mathbb{P}$, and the statement in the lemma follows. \square

The next result shows that F contains elements of a certain type.

Lemma 3.4. *For any integer m and any $\varepsilon > 0$ there exists an element $x^t u$ in F with $u \in \mathbb{U}$ and $t \in (\pi m, \pi m + \varepsilon)$.*

Proof. Let h_1 be the element $w_2 w_1^{-1} w_2$ in F . Then, by Lemma 3.3, we have $V_{h_1^{-1}}(x^{\beta+\pi}) = x^0$ where $\beta = \arg(\frac{3}{4})$. It follows that $V_{h_1^{-1}}(x^0) < V_{h_1^{-1}}(x^\beta) = x^{-\pi}$ and that $V_{h_1^{-N}}(x^0) < x^{-\pi N}$ for any natural number N .

We conclude that for the given integer m there exists a natural number N and an integer $M < m$ with

$$V_{h_1^{-N}}(x^0) \in [x^{\pi M}, x^{\pi(M+1)}).$$

For ε the given real number, there exists by Lemma 3.3 and the continuity of V_g a δ with $0 < \delta < \varepsilon$ and

$$V_{h_1}([x^0, x^\delta]) \subseteq (x^\pi, x^{\pi+\frac{\pi}{2}})$$

and hence

$$(*) \quad V_{h_1}([x^{\pi k}, x^{\pi k+\delta}]) \subseteq (x^{\pi(k+1)}, x^{\pi(k+1)+\frac{\pi}{2}})$$

follows for all $k \in \mathbb{Z}$.

By Lemma 3.1 there exists a natural number n_1 with

$$V_{w_1^{n_1} h}(x^0) \in [x^{\pi M}, x^{\pi M+\delta}] \quad \text{and} \quad h = h_1^{-N} \in F.$$

Hence, by $(*)$ we obtain

$$V_{h_1 w_1^{n_1} h}(x^0) \in (x^{\pi(M+1)}, x^{\pi(M+1)+\frac{\pi}{2}}).$$

By another application of Lemma 3.1, there exists a natural number n_2 with

$$V_{w_1^{n_2} h_1 w_1^{n_1} h}(x^0) \in (x^{\pi(M+1)}, x^{\pi(M+1)+\delta}) \subseteq (x^{\pi(M+1)}, x^{\pi(M+1)+\varepsilon}).$$

By repeating the last two steps $m - (M + 1)$ times, the statement of the lemma follows. \square

The next result shows that the cones P_H are indeed exceptional.

Proposition 3.5. *Let $H \supseteq F$ be a subgroup of \mathbb{G} and $P_H = \mathbb{P} \cap H$. Then:*

- a): P_H is an exceptional rank one cone in H .
- b): The mapping $\varphi : \mathcal{M}(\mathbb{P}) \rightarrow \mathcal{M}(P_H)$ with $\varphi(x^{\pi m} \mathbb{P}) = x^{\pi m} \mathbb{P} \cap H$, $m \in \mathbb{Z}$, defines an isomorphism between $\mathcal{M}(\mathbb{P})$ and $\mathcal{M}(P_H)$. The inverse of φ is given by $\varphi^{-1}(C) = \widehat{C} \mathbb{P}$ for C a divisorial P_H -ideal.

Proof. We recall that $\mathcal{M}(\mathbb{P})$ is the group of divisorial \mathbb{P} -ideals in \mathbb{G} (Definition 1.6) and that $\mathcal{M}(\mathbb{P}) = \text{gr}\{Q\} = \text{gr}(x^\pi \mathbb{P})$ by Theorems 1.9 and 3.2.

If C is a divisorial P_H -ideal in H , then $C\mathbb{P}$ is a \mathbb{P} -ideal in \mathbb{G} by Lemma 3.4. The divisorial closure $\widehat{C\mathbb{P}}$ of $C\mathbb{P}$ is therefore equal to some power of $x^\pi \mathbb{P}$ and $\widehat{C\mathbb{P}} = x^{\pi m} \mathbb{P}$ follows for some m in \mathbb{Z} . We want to prove that $\widehat{C\mathbb{P}} \cap H = C$ and assume that $hP_H \supseteq C$ for some $h \in H$. Then $hP_H \mathbb{P} = h\mathbb{P} \supseteq C\mathbb{P}$; hence $h\mathbb{P} \supseteq \widehat{C\mathbb{P}}$. Therefore, $hP_H = h\mathbb{P} \cap H \supseteq \widehat{C\mathbb{P}} \cap H$. It follows that $C = \widehat{C} \supseteq \widehat{C\mathbb{P}} \cap H \supseteq C$ and $C = \widehat{C\mathbb{P}} \cap H$. This shows that C being a divisorial P_H -ideal implies $C = x^{\pi m} \mathbb{P} \cap H$ for some m . We want to show next that $(x^{\pi n} \mathbb{P} \cap H) = x^{\pi n} \mathbb{P} \cap H$ for any n . Since $(x^{\pi n} \mathbb{P} \cap H)$ is divisorial, we know that $(x^{\pi n} \mathbb{P} \cap H) = x^{\pi m} \mathbb{P} \cap H$ for some m by the above argument.

By Lemma 3.5 there exist elements $x^{t_1} u_1, x^{t_2} u_2 \in F \subseteq H$ with $t_1 < t_2 \in \mathbb{R}$, $u_1, u_2 \in \mathbb{U}$ and $t_1, t_2 \in (\pi(n-1), \pi(n-1) + \frac{\pi}{2})$.

It follows that

$$x^{t_1} u_1 P_H = x^{t_1} u_1 \mathbb{P} \cap H \supseteq x^{t_2} u_2 \mathbb{P} \cap H = x^{t_2} u_2 P_H \supseteq x^{\pi n} \mathbb{P} \cap H.$$

Hence, $x^{\pi(n-1)} \mathbb{P} \cap H \supseteq (x^{\pi n} \mathbb{P} \cap H) \supseteq x^{\pi n} \mathbb{P} \cap H$.

If $x^{\pi(n-1)} \mathbb{P} \cap H = (x^{\pi n} \mathbb{P} \cap H)$, then this ideal would also be equal to $x^{t_1} u_1 P_H$ and $x^{t_2} u_2 P_H$. This would imply $x^{t_1} u_1 P_H \mathbb{P} = x^{t_1} \mathbb{P} = x^{t_2} u_2 P_H \mathbb{P} = x^{t_2} \mathbb{P}$, a contradiction that shows that $(x^{\pi n} \mathbb{P} \cap H) = (x^{\pi n} \mathbb{P} \cap H)$ for all n . This set of divisorial P_H -ideals does not contain $J(P_H)$, does not contain a completely prime ideal (Lemmas 3.3 and 3.5) and no ideal of the form $aJ(P_H) \neq J(P_H)$, $a \in P_H$, is completely prime in P_H . This shows that P_H has rank one and that $\mathcal{M}(P_H)$ is infinite cyclic with $Q_H = x^\pi \mathbb{P} \cap H$ as the positive generator of $\mathcal{M}(P_H)$. Since $J(P_H) \supset Q_H$, it follows from Theorem 1.9 that P_H is an exceptional rank one cone in H . This proves all statements in the lemma. \square

We consider now a condition that will guarantee that P_H is exceptional of type (C_k) .

Proposition 3.6. *Let H be a subgroup of \mathbb{G} containing F with $H \cap (\text{gr}\{x^\pi\} \times \mathbb{U}) = \text{gr}\{x^{\pi k}\} \times U(P_H)$ for some integer $k \geq 0$. Then the exceptional cone P_H has type (C_k) .*

Proof. It was shown in the previous proposition that P_H is an exceptional cone with $\mathcal{M}(P_H) = \text{gr}\{(x^\pi \mathbb{P} \cap H)\}$. To prove the statement in this proposition it must be shown that $\mathcal{H}(P_H) = \text{gr}\{x^{\pi k} P_H\}$, see Theorem 1.9. Hence, let $gP_H = P_H g$ be a principal ideal in H (see property e) in Section 1.1).

Then $g\mathbb{P} = P_H g\mathbb{P} = \mathbb{P} g\mathbb{P}$ by Lemma 3.4 and $g\mathbb{P} = \mathbb{P} g$ since \mathbb{P} has rank one. By Theorem 3.2, c) it follows that $g = x^{\pi m} u \in H$ for some integer m and $u \in \mathbb{U}$ and $g = x^{\pi k n} u$ for $u \in U(P_H)$ and some integer n by assumption. Therefore, $gP_H = x^{\pi k n} P_H$ and $\mathcal{H}(P_H) = \text{gr}\{x^{\pi k} P_H\} = \text{gr}\{\widehat{Q^k}\}$ follows for $Q = x^\pi \mathbb{P} \cap H$; P_H is exceptional of type (C_k) . \square

Theorem 3.7. *Let $H_k = \text{gr}\{w_1, w_2, x^{\pi k}\}$ be the subgroup of \mathbb{G} generated by F and the central element $x^{\pi k}$ for an integer $k \geq 0$. Then $P_k = \mathbb{P} \cap H_k$ is an exceptional rank one cone in H_k of type (C_k) .*

Proof. It is sufficient to verify the conditions in Proposition 3.7 for H_k .

Assume that

$$(*) \quad x^{\pi kp} w_1^{\nu_1} w_2^{\mu_1} w_1^{\nu_2} w_2^{\mu_2} \cdots w_1^{\nu_n} w_2^{\mu_n} = x^{\pi m} u \in H_k \cap (\text{gr} \{x^\pi\} \times \mathbb{U})$$

for some integers p, ν_i, μ_i for $i = 1, \dots, n$, and $u \in \mathbb{U}$. We apply the mapping τ (Theorem 2.5b)) to both sides of the above equation and obtain

$$(**) \quad (-1)^{kp} \begin{pmatrix} 1 & 2\nu_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\mu_1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 2\nu_n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\mu_n & 1 \end{pmatrix} = (-1)^m \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$$

where $u = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$ with $b, 0 < a \in \mathbb{R}$.

Since the entries of the matrices at the left side are all integers, it follows that a and a^{-1} are integers greater than zero; hence $a = a^{-1} = 1$. By a similar argument it follows that b is an even integer, $b = 2s$ for some s in \mathbb{Z} and $u = \begin{pmatrix} 1 & 2s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^s = w_1^s \in \tau(F)$ follows.

If $(-1)^{kp}(-1)^m = -1$, then it follows from $(**)$ that

$$-\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2\nu_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\mu_1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 2\nu_n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\mu_n & 1 \end{pmatrix} \begin{pmatrix} 1 & -b \\ 0 & 1 \end{pmatrix} \in \tau(F),$$

which is a contradiction, since the group $\tau(F)$ freely generated by $\tau(w_1)$ and $\tau(w_2)$ (see the remarks before Lemma 3.3) does not contain a nontrivial central element.

Therefore, $(-1)^{kp} = (-1)^m$ can be cancelled in $(**)$ and, using again the fact that $\tau(w_1) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\tau(w_2) = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ are free generators of $\tau(F)$, it follows that $n = 1$, $b = 2\nu_1$ if we ignore exponents that could be zero. With $u = w_1^{\nu_1}$ we can rewrite $(*)$ as: $x^{\pi kp} w_1^{\nu_1} = x^{\pi m} w_1^{\nu_1}$. It follows that $\nu_1 = s$, $m = kp$ and $u = w_1^s \in U(P_k)$; the condition in Proposition 3.7 is satisfied and Theorem 3.8 follows. \square

4. EXAMPLES OF EXCEPTIONAL RANK ONE CHAIN DOMAINS

In this section we construct domains S_k associated with the exceptional cones P_k of type (C_k) as described in Theorem 3.8.

In Lemma 2.6(c) it was proved that $T\mathbb{U}$ is an Ore domain for any skew field T and the subgroup \mathbb{U} of \mathbb{G} . We denote by K the skew field of quotients of $T\mathbb{U}$ for a given skew field T ; for example, $T = \mathbb{Q}$, the rationals. Let $K\{\mathbb{G}\}$ be the right K -vector space and left T -vector space consisting of all series

$$\gamma = x^{t_1} k_1 + x^{t_2} k_2 + \cdots$$

with $t_1 < t_2 < \dots$, $k_i \in K$, and $\text{supp}(\gamma) = \{x^{t_i} \mid k_i \neq 0\}$ well ordered.

We call $\text{supp}(\gamma)$ the support of the series γ . If $k_1 \neq 0$, then $v(\gamma) = x^{t_1} \in R$ is the norm of γ and $v(0) = \infty$ for $\gamma = 0$.

Let $Q = \text{End } K\{\mathbb{G}\}_K$ be the endomorphism ring of the K -vector space $K\{\mathbb{G}\}_K$. For $q \in Q$ and $\gamma \in K\{\mathbb{G}\}$ we write $q[\gamma]$ for the image of γ under q . The representation $V : \mathbb{G} \rightarrow \text{Aut}(R, \geq)$ considered in Section 2.3 can be extended to a mapping V defined on Q with

$$V_q(x^t) = v(q[x^t]), \quad V_q(\infty) = \infty$$

for $q \in Q$, $x^t \in R$, and $V_q : (R, \infty) \rightarrow (R, \infty)$. It follows that

$$V_{a+b}(x^t) \geq \min \{V_a(x^t), V_b(x^t)\}$$

for any $a, b \in Q$ and $x^t \in R$. However, V_{ab} is not equal to $V_a \circ V_b$ in general.

We recall a definition and a result given by Mathiak in [17].

Definition 4.1. Let D be a skew field and (Γ, \leq) a linearly ordered set. Then a mapping $V : D^* \rightarrow \text{Aut}(\Gamma, \leq)$ is called an M -valuation if the following conditions hold:

MV1. $V_{ab} = V_a \circ V_b$ for any $a, b \in D^*$;

MV2. $V_{a+b}(h) \geq \min\{V_a(h), V_b(h)\}$ for any $a, b \in D^*$ with $a + b \neq 0$ and $h \in \Gamma$.

If we add the symbol ∞ for infinity to Γ and define $V_0(h) = \infty$ and $V_a(\infty) = \infty$ for all $h \in \Gamma$, $0, a \in D$, then MV1 and MV2 will be valid for all elements $a, b \in D$ and all $h \in \Gamma \cup \{\infty\}$.

The next result follows almost directly from the previous definition; see also [16] and [17].

Proposition 4.2. *Let $V : D^* \rightarrow \text{Aut}(\Gamma, \leq)$ be an M -valuation for a skew field D and a linearly ordered set (Γ, \leq) and let h be an element in Γ .*

Then the set $S_h = \{d \in D \mid V_d(h) \geq h\}$ is a total subring of D . Conversely, any total subring S in a skew field D can be obtained in this way for $\Gamma = \{aS \mid a \in D^\}$, $aS \geq bS$ if and only if $aS \subseteq bS$ and $V_d(aS) = daS$. The ring S coincides with S_h for $h = S \in \Gamma$. \square*

The space $K\{\mathbb{G}\}$ introduced above is also a left \mathbb{G} -module if we define for $g \in \mathbb{G}$ and $\gamma = \sum x^{t_i} k_i \in K\{\mathbb{G}\}$ that

$$g\gamma = x^{t'_1}(u_1 k_1) + x^{t'_2}(u_2 k_2) + x^{t'_3}(u_3 k_3) + \dots$$

where $g \cdot x^{t_i} = x^{t'_i} u_i$ for $u_i \in \mathbb{U} \subseteq K$, $t'_i \in \mathbb{R}$. It follows from Lemma 2.6(d) that $t'_1 < t'_2 < t'_3 < \dots$ is also well ordered and hence $g\gamma \in K\{\mathbb{G}\}$. The group ring $T\mathbb{G}$ can therefore be considered as a subring of Q .

If A is any subring of Q , then we define $\mathcal{D}[0, A] = A$ and $\mathcal{D}[n+1, A]$ as the subring of Q generated by $\mathcal{D}[n, A]$ and all inverses of elements of $\mathcal{D}[n, A]$ in Q . The union

$$\bigcup_{n=0}^{\infty} \mathcal{D}[n, A] = \mathcal{D}[A]$$

is called the *rational closure* of A in Q . Let $\mathbb{D} = \mathcal{D}[T\mathbb{G}]$ be the rational closure of the group ring $T\mathbb{G}$ in Q .

The following result can be found in [11] (see [10] also):

Theorem 4.3. a) *The rational closure \mathbb{D} of $T\mathbb{G}$ in Q is a skew field.*

b) *The mapping V restricted to \mathbb{D}^* is an M -valuation of \mathbb{D}^* to $\text{Aut}(\mathbb{D}, \leq)$.*

c) *The ring $S = \{d \in \mathbb{D} \mid V_d(x^0) \geq x^0\}$ is an exceptional rank one chain order in \mathbb{D} of type (C_1) associated with the exceptional cone \mathbb{P} in the group \mathbb{G} . \square*

In order to construct skew fields that contain rank one exceptional chain orders of type (C_k) we consider the rational closure $D_k = \mathcal{D}[TH_k]$ of the group ring TH_k for the group $H_k = \text{gr}\{w_1, w_2, x^{\pi k}\}$ (see Theorem 3.8) in $Q = \text{End } K\{\mathbb{G}\}_K$.

Since $D_k \subseteq \mathbb{D} = \mathcal{D}[T\mathbb{G}] \subset Q$ and \mathbb{D} is a skew field by the above theorem, it follows that D_k is also a skew field and $S_k = S \cap D_k$ is a total subring of D_k .

It follows from Corollary 1.10 and Theorem 3.8 that S_k is an exceptional rank one chain domain of type (C_k) if the following theorem is proved:

Theorem 4.4. *The total subring $S_k = S \cap D_k$ is associated with the cone $P_k = \mathbb{P} \cap H_k$.*

Before this theorem can be proved, we need the result in the following lemma.

Lemma 4.5. *Let $\gamma \in K\{\mathbb{G}\}$. Then*

$$(*) \quad \bigcup_{d \in D_k} \text{supp } d[\gamma] \subseteq \bigcup_{g \in H_k} V_g(\text{supp } \gamma).$$

Proof. Let Y_γ be the right side in (*). Then in order to prove (*) it is sufficient to prove

$$(**) \quad \text{supp } d[\gamma] \subseteq Y_\gamma$$

for any $\gamma \in K\{\mathbb{G}\}$ and any $d \in D_k = \bigcup \mathcal{D}[n, TH_k]$. We will prove this in five steps using induction on n for n the smallest index with $d \in \mathcal{D}[n, TH_k]$.

STEP 1. Assume that $d = x^t u \in H_k$, $u \in \mathbb{U}$ and that $\gamma = \sum_{i < \Lambda} x^{t_i} k_i \in K\{\mathbb{G}\}$, $0 \neq k_i \in K$ for all ordinals $i < \Lambda$.

Then $d\gamma = \sum_{i < \Lambda} x^{t_i} (u_i k_i)$ for $x^t u x^{t_i} = x^{t_i} u_i \in \mathbb{G}$, $u_i \in \mathbb{U}$. Hence, $\text{supp } d[\gamma] = \text{supp } d\gamma = \{x^{t_i} \mid i < \Lambda\} = \{V_d(x^{t_i}) \mid i < \Lambda\} \subseteq \{V_g(x^{t_i}) \mid g \in H_k, i < \Lambda\} = Y_\gamma$.

STEP 2. The inclusion (**) follows immediately for $d \in T$.

STEP 3. Assume that $a, b \in D_k$ with $\text{supp } a[\gamma] \cup \text{supp } b[\gamma] \subseteq Y_\gamma$ for any $\gamma \in K\{\mathbb{G}\}$. Then $\text{supp } (a+b)[\gamma] \subseteq \text{supp } a[\gamma] \cup \text{supp } b[\gamma] \subseteq Y_\gamma$. Further, $V_g(Y_\gamma) = Y_\gamma$ for $g \in H_k$ and hence

$$\text{supp } (a \cdot b)[\gamma] \subseteq Y_{b[\gamma]} = \bigcup_{g \in H_k} V_g(\text{supp } b[\gamma]) \subseteq \bigcup_{g \in H_k} V_g(Y_\gamma) = Y_\gamma.$$

STEP 4. It follows from Steps 1-3 that the statement (**) is true for all $d \in TH_k = \mathcal{D}[0, TH_k]$ and any $\gamma \in K\{\mathbb{G}\}$.

STEP 5. Assume that (**) is true for elements $d \in \mathcal{D}[n-1, TH_k]$ for some $n \geq 1$ and all $\gamma \in K\{\mathbb{G}\}$.

Let $d = p^{-1} \in \mathcal{D}[n, TH_k]$ with $p \in \mathcal{D}[n-1, TH_k]$. We consider $\beta = d[\gamma]$ and decompose β into the sum $\beta = \beta_0 + \beta_1$ with $\text{supp } (\beta_0) \subseteq Y_\gamma$ and $\text{supp } (\beta_1) \cap Y_\gamma = \emptyset$.

Then $\gamma = p[\beta] = p[\beta_0] + p[\beta_1]$.

By the induction hypothesis, it follows that

$$\text{supp } (p[\beta_0]) \subseteq \bigcup_{g \in H_k} V_g(\text{supp } \beta_0) \subseteq \bigcup_{g \in H_k} V_g(Y_\gamma) \subseteq Y_\gamma.$$

Hence, $\text{supp } (p[\beta_1]) = \text{supp } (\gamma - p[\beta_0]) \subseteq \text{supp } \gamma \cup \text{supp } (p[\beta_0]) \subseteq Y_\gamma$. On the other hand, $\text{supp } (p[\beta_1]) \subseteq Y_{\beta_1}$ since $p \in \mathcal{D}[n-1, TH_k]$. If we assume that there exists an element h in $\text{supp } (p[\beta_1])$, then, on the one hand,

$$h = V_g h' \quad \text{for some } g \in H_k \quad \text{and some } h' \in \text{supp } (\gamma)$$

and on the other hand,

$$h = V_{g'}(h'') \quad \text{for some } g' \in H_k \quad \text{and some } h'' \in \text{supp } (\beta_1).$$

This implies $h'' = V_{(g')^{-1}g}(h') \in Y_\gamma \cap \text{supp } (\beta_1) = \emptyset$, a contradiction that shows that $\text{supp } (\beta_1)$ is empty and $\text{supp } (\beta) = \text{supp } (\beta_0) \subseteq Y_\gamma$.

The ring $\mathcal{D}[n, TH_k]$ is generated by $\mathcal{D}[n-1, TH_k]$ and all elements p^{-1} for $p \in \mathcal{D}[n-1, TH_k] \setminus \{0\}$, and it now follows by an application of Step 3 that (**) is true for all elements in $\mathcal{D}[n, TH_k]$ which completes the induction and proves the lemma (see also: [11]). \square

We now return to the proof of Theorem 4.4.

Let d be a nonzero element in D_k . Since $D_k \subseteq \mathbb{D}$ and S is associated with the cone \mathbb{P} , the element d can be decomposed as follows:

$$d = x^t m = q x^{t'} \quad \text{with } m, q \in U(S), \mathbb{P} x^t \mathbb{P} = \mathbb{P} x^{t'} \mathbb{P}$$

(see Definition 1.2). It follows from (*) in Lemma 4.5 with $\gamma = x^0$ that

$$\text{supp } d[x^0] \subseteq \bigcup_{g \in H_k} V_g(x^0) = \bigcup_{g \in H_k} v(g).$$

Hence, $v(d[x^0]) = V_d(x^0) = V_{x^t} \circ V_m(x^0) = V_{x^t}(x^0) = x^t$ since $m \in U(S)$, and hence $x^t = v(g)$, $g = x^t u$, $u \in \mathbb{U}$ for some element $g \in H_k$.

It follows that $d = (x^t u)(u^{-1}m)$ for $x^t u \in H_k$ and $u^{-1}m = (x^t u)^{-1}d \in D_k$. Further, $u^{-1}m \in \mathbb{U} \cdot U(S) \cap D_k = U(S_k)$, since $\mathbb{U} \subseteq U(S)$ and $S \cap D_k = S_k$.

Applying the same arguments to the element $d^{-1} = x^{-t'} q^{-1}$, we conclude that there exists an element $g' \in H_k$ with $g' = x^{-t'} w$ for some $w \in \mathbb{U}$. Hence, we obtain a decomposition

$$d^{-1} = (x^{-t'} w)(w^{-1} q^{-1}) \quad \text{for} \quad w^{-1} q^{-1} = (g')^{-1} d^{-1} \in D_k \cap U(S) = U(S_k).$$

This proves the first half of condition (ii) in Definition 1.2, if we write $d = (qw)(w^{-1}x^{t'})$, $qw \in U(S_k)$, $w^{-1}x^{t'} = (g')^{-1} \in H_k$. It remains to prove the equality

$$P_k x^t u P_k = P_k w^{-1} x^{t'} P_k.$$

Let $w^{-1}x^{t'} = x^{t''} u''$ for some $u'' \in \mathbb{U}$ and $t'' \in \mathbb{R}$. Since S is associated with \mathbb{P} , it follows that

$$\mathbb{P} x^{t'} \mathbb{P} = \mathbb{P} x^{t''} \mathbb{P}.$$

Therefore,

$$\begin{aligned} P_k x^t u P_k &= P_k x^t u (\mathbb{P} \cap H_k) = P_k x^t \mathbb{P} \cap H_k \\ &= \mathbb{P} x^t \mathbb{P} \cap H_k = \mathbb{P} x^{t'} \mathbb{P} \cap H_k = \mathbb{P} x^{t''} \mathbb{P} \cap H_k \\ &= P_k x^{t''} u'' \mathbb{P} \cap H_k = P_k x^{t''} u'' P_k \\ &= P_k w^{-1} x^{t'} P_k \end{aligned}$$

where we used Lemma 3.4 for the third and sixth equality.

This completes the proof of Theorem 4.4. □

Corollary 4.6. *The chain domain S_k is exceptional of rank one and of type (C_k) .*

REFERENCES

1. G. M. Bergman. *Right-orderable groups that are not locally indicable*. Pacific J. Math. **147** (1991), 243-248. MR **92e**:20030
2. H. H. Brungs, H. Marubayashi, and E. Osmanagic. *A classification of prime segments in simple Artinian rings*. Proc. Amer. Math. Soc. **128** (2000), 3167-3175. MR **2001b**:16055
3. H. H. Brungs and M. Schröder. *Prime segments of skew fields*, Canad. J. Math. **47** (1995), 1148-1167. MR **97c**:16021
4. H. H. Brungs and G. Törner. *Chain rings and prime ideals*. Arch. Math. **27** (1976), 253-260. MR **54**:7537
5. H. H. Brungs and G. Törner. *Extensions of chain rings*. Math. Zeit. **185** (1984), 93-104. MR **85d**:16012
6. H. H. Brungs and G. Törner. *Ideal theory of right cones and associated rings*. J. Algebra **210** (1998), 145-164. MR **99k**:20113
7. P. M. Cohn, Skew Fields, Cambridge University Press, 1995. MR **97d**:12003
8. N. I. Dubrovin. *Chain domains*. Moscow Univ. Math. Bull. Ser. 1. **37**(1980), 51-54. MR **81g**:16004
9. N. I. Dubrovin. *An example of a chain prime ring with nilpotent elements*. Mat. Sbornik **48** (1984), 437-444. MR **84f**:16012
10. N. I. Dubrovin. *The rational closure of group rings of left-orderable groups*. Mat. Sbornik **184**(7) (1993), 3-48. MR **94g**:16035

11. N. I. Dubrovin. *The rational closure of group rings of left-ordered groups*. SM-DU-254, Duisburg, 1994, 96 pp.
12. T. V. Dubrovin and N. I. Dubrovin. *Cones in groups*. Sbornik Mathematics **187**(7) 1996, 1005-1019. MR **98c**:20082
13. L. Fuchs. *Teilweise geordnete algebraische Strukturen*, Vandenhoeck und Ruprecht, Göttingen, 1966. MR **34**:4386
14. V. K. Goel and S. K. Jain. π -injective modules and rings whose cyclics are π -injective. Comm. Algebra **6** (1978), 59-73. MR **58**:11016
15. M. I. Kargapolov and Ju. I. Merzljakov. *Fundamentals of the Theory of Groups*. Springer-Verlag, 1979. MR **80k**:20002
16. K. Mathiak. *Zur Bewertungstheorie nicht kommutativer Körper*, J. Algebra **73**, (1981), No. 2, 586-600. MR **83c**:12026
17. K. Mathiak. *Valuations of skew fields and projective Hjelmslev spaces*. Lecture Notes in Math. 1175, Springer-Verlag, 1986. MR **87g**:16002
18. B. L. Osofsky. *Noncommutative rings whose cyclic modules have cyclic injective hulls*. Pacific J. Math. **25** (1968), 331-340. MR **38**:186
19. E. C. Posner. *Left valuation rings and simple radical rings*. Trans. Amer. Math. Soc. **107** (1963), 458-465. MR **27**:3665

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF ALBERTA, EDMONTON T6G 2G1, CANADA

E-mail address: hbrungs@math.ualberta.ca

DEPARTMENT OF MATHEMATICS, VLADIMIR STATE UNIVERSITY, GORKI STR. 87, 600026 VLADIMIR, RUSSIA

E-mail address: ndubrovin@mail.ru

ON THE SPECTRAL SEQUENCE CONSTRUCTORS OF GUICHARDET AND STEFAN

DONALD W. BARNES

ABSTRACT. The concept of a spectral sequence constructor is generalised to Hopf Galois extensions. The spectral sequence constructions that are given by Guichardet for crossed product algebras are also generalised and shown to provide examples. It is shown that all spectral sequence constructors for Hopf Galois extensions construct the same spectral sequence.

1. INTRODUCTION

A. Guichardet [3] has given two constructions for a spectral sequence with $E_2^{p,q} = H^p(G, H^q(B, X))$ and target the Hochschild cohomology $H^\bullet(G \times_\alpha B, X)$ of the crossed product algebra $A = G \times_\alpha B$, where G is a group, B is an algebra, α is a representation of G by automorphisms of B and the multiplication in A is given by

$$(g, b)(g', b') = (gg', \alpha_{g'}^{-1}(b)b').$$

Here, X is a left and right A -bimodule or, equivalently, a left module over $A^e = A \otimes A^{\text{op}}$ where A^{op} is the opposed algebra of A . These constructions are analogous to the Hochschild-Serre constructions for the spectral sequence of a group extension. He has asked if the methods of Barnes [1] can be used to show that they construct the same spectral sequence.

D. Stefan [5] has given a spectral sequence, based on the Grothendieck composite functor spectral sequence, for the cohomology of a Hopf Galois extension.

The three contexts have some features in common. All have a “large” algebra A and the category \mathcal{A} of A -(bi)modules, a subalgebra B , a “small” algebra C which plays the role of a quotient of A by B and the category \mathcal{C} of C -modules and a category \mathcal{D} in which the filtered cochain complexes are constructed. All have a left exact functor $\phi : \mathcal{A} \rightarrow \mathcal{C}$ and a left exact functor $\psi : \mathcal{C} \rightarrow \mathcal{D}$, and the spectral sequences have as target the right derived functors of the composite $\theta = \psi \circ \phi$. Throughout this paper, ϕ , ψ and θ will denote these functors.

In Barnes [1], A is an augmented algebra over a commutative ring \mathfrak{K} and C is the quotient $A//B$ of A by a normal augmented subalgebra. A , B and $A//B$ are all assumed to be projective as \mathfrak{K} -modules. The functor ϕ is given for the left A -module

Received by the editors April 30, 2001.

2000 *Mathematics Subject Classification.* Primary 18G40, 16W30; Secondary 16E40.

Key words and phrases. Spectral sequence, crossed product, comodule algebra, Hopf Galois extension.

This work was done while the author was an Honorary Associate of the School of Mathematics and Statistics, University of Sydney.

X by

$$\phi X = X^B = \{x \in X \mid bx = \epsilon(b)x \text{ for all } b \in B\}.$$

For a C -module Y , $\psi Y = Y^C$ and we have $\psi\phi X = X^A$ for any A -module X . In this context, ϕ has a left adjoint $j : \mathcal{C} \rightarrow \mathcal{A}$, and use is made of the counit $\pi = j\phi : \mathcal{A} \rightarrow \mathcal{A}$ of the adjunction. Note that $\phi j = \text{id} : \mathcal{C} \rightarrow \mathcal{C}$. We shall refer to this as the HS context.

In Guichardet's paper, A is the crossed product $G \times_\alpha B$, where B is an algebra over the field \mathfrak{K} , C is the group algebra $\mathfrak{K}G$ and the functor ϕ is given for the left A^e -module X by

$$\phi X = X^B = \{x \in X \mid bx = xb \text{ for all } b \in B\}$$

with the action of G given by $gx = (g, 1)x(g^{-1}, 1)$ for $g \in G$ and $x \in X^B$. For the $\mathfrak{K}G$ -module Y ,

$$\psi Y = Y^G = \{y \in Y \mid gy = y \text{ for all } g \in G\}.$$

In this context, ϕ does not have a left adjoint j . The assumption that \mathfrak{K} is a field can be weakened if in some places we replace "injective" by "relatively injective". We require that \mathfrak{K} is a commutative ring and that B is \mathfrak{K} -projective. We shall refer to this as the G context.

In Stefan [5], C is a Hopf algebra over the field \mathfrak{K} , A is a C -comodule algebra and B is the subalgebra of coinvariants. Thus we have an algebra morphism $\Delta_A : A \rightarrow A \otimes C$ making A a right C -comodule, and

$$B = A^{\text{co}C} = \{a \in A \mid \Delta_A(a) = a \otimes 1\}.$$

Stefan refers to this situation as "the extension A/B ". It is assumed to be C -Galois, which we explain in section 2 below. As in the G context, for the A -bimodule X , we set $\phi X = X^B$. This is made into a right C -module using the action defined by Stefan [5, Proposition 2.3] and explained following Lemma 2.1 below. For the C -module Y , we put $\psi Y = Y^C$. Again, we weaken the assumption that \mathfrak{K} is a field. We require that \mathfrak{K} is a commutative ring and that A , B and C are \mathfrak{K} -projective. We refer to this as the S context. It generalises the G context since the crossed product algebra $A = G \times_\alpha B$ becomes a C -comodule algebra if we set $\Delta_A(g, b) = (g, b) \otimes g$ for $g \in G$ and $b \in B$. If in the HS (Hochschild-Serre) context, A is a Hopf algebra, we may regard it as a C -comodule algebra with the comodule structure given by the comultiplication of A followed by the natural homomorphism $A \otimes A \rightarrow A \otimes A//B$. A left A -module X may be regarded as a bimodule by setting $xa = \epsilon(a)x$ for $a \in A$ and $x \in X$, where ϵ is the augmentation. This does not change $\phi X = X^B$, now defined as $\{x \in X \mid bx = xb \text{ for all } b \in B\}$, nor does it change the $R^q\phi(X)$, although it does change the injective modules used for their calculation.

2. PRELIMINARIES

We follow the notation for comodule algebras used in Schneider [4], with the exception that we denote the Hopf algebra by C , reserving the symbol H for cohomology. Thus we have the comodule structure map, $\Delta_A : A \rightarrow A \otimes C$, and express the image of an element $a \in A$ by $\Delta_A(a) = \sum a_0 \otimes a_1$. The comultiplication $\Delta_C : C \rightarrow C \otimes C$ is written $\Delta_C(c) = \sum c_1 \otimes c_2$. The augmentation of C is denoted by ϵ and the antipode by S . The canonical map $\text{can} : A \otimes_B A \rightarrow A \otimes C$ is defined by $\text{can}(a \otimes_B a') = \sum aa'_0 \otimes a'_1$. That A/B is a Hopf Galois extension means that can is invertible, which we always assume. Thus for $c \in C$, there exist elements

$r_i(c), l_i(c) \in A$, not uniquely determined, such that $\mathbf{can}^{-1}(1 \otimes c) = \sum r_i(c) \otimes_B l_i(c)$, which is unique. We shall need the identities proved in Schneider [4, Remark 3.4(2)]. Throughout his paper, Schneider assumes \mathfrak{K} to be a field, but his proof of the identities makes no use of that assumption. For the convenience of the reader, we list the identities here.

Lemma 2.1. *For all $a \in A$, $b \in B$ and $c, c' \in C$, the following identities hold:*

- (a) $\sum b r_i(c) \otimes_B l_i(c) = \sum r_i(c) \otimes_B l_i(c) b$,
- (b) $\sum a_0 r_i(a_1) \otimes_B l_i(a_1) = 1 \otimes_B a$,
- (c) $\sum r_i(c) l_i(c) = \epsilon(c)$,
- (d) $\sum r_i(c) \otimes_B l_i(c)_0 \otimes l_i(c)_1 = \sum r_i(c_1) \otimes_B l_i(c_1) \otimes c_2$,
- (e) $\sum r_i(c)_0 \otimes_B l_i(c) \otimes r_i(c)_1 = \sum r_i(c_2) \otimes_B l_i(c_2) \otimes S(c_1)$,
- (f) $\sum r_i(cc') \otimes_B l_i(cc') = \sum r_i(c') r_j(c) \otimes_B l_j(c) l_i(c')$,
- (g) $\sum r_i(c_1) \otimes_B l_i(c_1) r_j(c_2) \otimes_B l_j(c_2) = \sum r_i(c) \otimes_B 1 \otimes_B l_i(c)$.

Following Stefan [5, Proposition 2.3], we use the above relations to define a right C -module structure on X^B for any left A^e -module X . For $x \in X^B$ and $c \in C$ we put $x \cdot c = \sum r_i(c) x l_i(c)$. This is well defined since $\sum r_i(c) \otimes_B l_i(c)$ is a well-defined element of $A \otimes_B A$ and $bx = xb$ for all $b \in B$. From 2.1(a), it follows that $x \cdot c \in X^B$. If a left action of C on X^B is preferred, one may be defined by setting $c \cdot x = x \cdot (Sc)$. The assertion of Lemma 2.2 below holds for this left action provided that the antipode S is bijective.

Lemma 2.2. *For an A^e -module X , $(X^B)^C = X^A$.*

Proof. For $x \in X^A$ and $c \in C$, we have

$$x \cdot c = \sum r_i(c) x l_i(c) = \sum r_i(c) l_i(c) x = \epsilon(c) x$$

by 2.1(c). Thus $x \in (X^B)^C$. Conversely, if $x \in (X^B)^C$, then by 2.1(b),

$$xa = 1xa = \sum a_0 r_i(a_1) x l_i(a_1) = \sum a_0 (x \cdot a_1) = \sum a_0 \epsilon(a_1) x = ax$$

for all $a \in A$. Thus $x \in X^A$. □

For the crossed product algebra $A = G \rtimes_{\alpha} B$, the canonical map is given by

$$\mathbf{can}((g, b) \otimes_B (g', b')) = (g, b)(g', b') \otimes g' = (gg', \alpha_{g'}^{-1}(b)b') \otimes g'.$$

In particular, $\mathbf{can}((g^{-1}, 1) \otimes_B (g, 1)) = (1, 1) \otimes g$; so we can take $r(g) = (g^{-1}, 1)$ and $l(g) = (g, 1)$. The right action of $C = \mathfrak{K}G$ on X^B becomes $x \cdot g = (g^{-1}, 1)x(g, 1)$ for $x \in X$ and $g \in G$. Converting this to a left action gives $g \cdot x = (g, 1)x(g^{-1}, 1)$, which is the action used in the Guichardet paper [3].

Note that, in the G context, $A^e = A \otimes A^{\text{op}}$ is free as a right B^e -module. In the S context, we assume that A is flat as left and right B -module. It then follows that A^e is flat as right B^e -module. In the HS context, we assume that A is at least projective as right B -module. At some points in [1], the stronger assumption that the module quotient A/B is projective as right B -module is used. In all the contexts, by Barnes [1, Lemma I.4.3], every injective left A - or A^e -module is injective as B - or B^e -module. Further, every injective of \mathcal{A} or \mathcal{C} is injective as \mathfrak{K} -module.

For any \mathfrak{K} -module X , the A -module coinduced from X is the module $X^* = \text{Hom}_{\mathfrak{K}}(A, X)$ with the action $(af)(a') = f(a'a)$ for $f \in X^*$ and $a, a' \in A$. (For the coinduced right module, the action is given by $(fa)(a') = f(aa')$.) If X is itself a left A -module, then the map $\sigma : X \rightarrow X^*$ defined by $(\sigma x)(a) = ax$ for $x \in X$ and

$a \in A$ is a \mathfrak{K} -split A -module monomorphism. X^* is a relatively injective left A -module. (See for example, Barnes [1, Lemma II.2.4, p. 20].) A module is relatively injective if and only if it is a direct summand of a coinduced module. By Barnes [1, Lemma II.3.9, p. 28], if A is right B -projective, then every relatively injective left A -module is relatively injective as left B -module. Thus in the G context, every relatively injective A^e -module is relatively injective as B^e -module. In the S context, to get this conclusion, we must strengthen Stefan's assumption that A is left and right B -flat to A being left and right B -projective, although as the next lemma shows, this strengthening is unnecessary if, as in [5], it is assumed that \mathfrak{K} is a field, since then, every module is \mathfrak{K} -injective.

Lemma 2.3. *If X is \mathfrak{K} -injective, then the coinduced module X^* is injective.*

Proof. Let $i : V \rightarrow W$ be a monomorphism and $\alpha : V \rightarrow X^*$ a homomorphism of A -modules.

$$\begin{array}{ccccc} 0 & \longrightarrow & V & \xrightarrow{i} & W \\ & & \downarrow \alpha & & \\ & & X^* & & \end{array}$$

We want to construct a homomorphism $\beta : W \rightarrow X^*$ such that $\beta i = \alpha$. For $v \in V$, we have $\alpha v \in \text{Hom}_{\mathfrak{K}}(A, X)$; so for $a \in A$, we have $(\alpha v)(a) \in X$. We may regard α as a function $A \times V \rightarrow X$, writing $\alpha(a, v)$ for $(\alpha v)(a)$. For $b \in A$, we have $\alpha(bv) = b(\alpha v)$. So $(\alpha(bv))(a) = (\alpha v)(ab)$; that is, $\alpha(ab, v) = \alpha(a, bv)$, and in particular $\alpha(a, v) = \alpha(1, av)$. Putting $\bar{\alpha}(v) = (\alpha v)(1)$ creates the diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & V & \xrightarrow{i} & W \\ & & \downarrow \bar{\alpha} & & \\ & & X & & \end{array}$$

of \mathfrak{K} -modules. Since X is \mathfrak{K} -injective, there exists a \mathfrak{K} -homomorphism $\bar{\beta} : W \rightarrow X$ such that $\bar{\beta} i = \bar{\alpha}$. Define $\beta : W \rightarrow X^*$ by $\beta(w)(a) = \bar{\beta}(aw)$. Then for $b \in A$, we have

$$\beta(bw)(a) = \bar{\beta}(abw) = (\beta w)(ab) = (b\beta(w))(a);$$

so β is an A -module homomorphism. We have

$$(\beta i v)(a) = \bar{\beta}(a(i v)) = \bar{\beta}(i(av)) = \bar{\alpha}(av) = (\alpha v)(a)$$

and $\beta i = \alpha$. □

We have defined a right C -module action on X^B for any left A^e -module X . We need another description of that action in the case where X is a coinduced module.

Lemma 2.4. *Let V be a \mathfrak{K} -module and let $V^* = \text{Hom}(A^e, V)$ be the coinduced A^e -module. Then V^{*B} is isomorphic to the coinduced C -module $\text{Hom}(C, \text{Hom}(A, V))$.*

Proof. The action of $a \otimes a' \in A^e$ on $f \in V^*$ is given by

$$((a \otimes a')f)(x \otimes y) = f((x \otimes y)(a \otimes a')) = f(xa \otimes a'y)$$

for $x, y \in A$. Now

$$\begin{aligned} V^{*B} &= \{f \in V^* \mid (b \otimes 1)f = (1 \otimes b)f \text{ for all } b \in B\} \\ &= \{f \in V^* \mid f(xb \otimes y) = f(x \otimes by) \text{ for all } b \in B \text{ and } x, y \text{ in } A\}. \end{aligned}$$

Thus V^{*B} can be identified with $\text{Hom}(A \otimes_B A, V)$, and so, using the canonical map, with $\text{Hom}(A \otimes C, V) = \text{Hom}(C, \text{Hom}(A, V))$. We calculate the right C -module action on $\text{Hom}(C, \text{Hom}(A, V))$ induced by these identifications. For $f \in V^{*B}$ and $c \in C$, from the A^e -action on V , we get $fc = \sum (r_i(c) \otimes l_i(c))f$. Thus

$$(fc)(x \otimes y) = \sum f(xr_i(c) \otimes l_i(c)y).$$

For $f \in \text{Hom}(C, \text{Hom}(A, V))$ we have the corresponding $f' \in \text{Hom}(A \otimes C, V)$ given by $f'(a \otimes c) = f(c)(a)$ and $f'' \in \text{Hom}(A \otimes_B A, V)$ given by

$$f''(x \otimes_B y) = f'(\text{can}(x \otimes_B y)) = \sum f'(xy_0 \otimes y_1).$$

Thus,

$$f(c)(a) = f'(a \otimes c) = \sum f''(ar_i(c) \otimes_B l_i(c)).$$

So for $c' \in C$, fc' is given by

$$\begin{aligned} (fc')(c)(a) &= \sum (f''c')(ar_i(c) \otimes_B l_i(c)) \\ &= \sum f''(ar_i(c)r_j(c') \otimes_B l_j(c')l_i(c)) \\ &= \sum f''(ar_i(c'c) \otimes_B l_i(c'c)) \quad \text{by 2.1(f)} \\ &= f(c'c)(a). \end{aligned}$$

Thus the action $(fc')(c) = f(c'c)$ is that of the coinduced right C -module. \square

The next result strengthens Stefan [5, Proposition 3.2]. The corresponding result in the HS context follows easily from the fact that every $A//B$ -module is an A -module and that Q^B is a submodule of Q .

Lemma 2.5 . *Let Q be a relatively injective A^e -module. Then Q^B is a relatively injective C -module. If Q is injective, then Q^B is injective.*

Proof. Q is a direct summand of some coinduced module $V^* = \text{Hom}_{\mathfrak{K}}(A^e, V)$. So to prove Q^B relatively injective, it is sufficient to show that $\text{Hom}_{\mathfrak{K}}(A^e, V)^B$ is relatively injective. But $\text{Hom}(C, \text{Hom}(A, V))$ is relatively injective; so by Lemma 2.4, V^{*B} is relatively injective. Thus Q^B is relatively injective. If Q is injective, we may take $V = Q$. Since Q is \mathfrak{K} -injective, by Lemma 2.3, $\text{Hom}(A, Q)$ is also \mathfrak{K} -injective and $\text{Hom}(C, \text{Hom}(A, Q))$ is an injective C -module. Thus Q^B is injective. \square

3. SPECTRAL SEQUENCE CONSTRUCTORS GENERALISED

In Barnes [1, Chapter III] in the HS context, a spectral sequence constructor for (ϕ, ψ) was defined to be a functor F from \mathcal{A} to filtered cochain complexes in \mathcal{D} such that

- (1) F is exact (in every filtration).
- (2) F is acyclic on injectives; that is, if Q is injective, then $H^n(FQ) = 0$ for $n > 0$ and for all p , $H^q(\bullet F^p Q / \bullet F^{p+1} Q) = 0$ for $q > 0$.
- (3) $E_1^{\bullet 0} F$ is exact on \mathcal{C} .
- (4) The inclusion $i : X^B \rightarrow X$ induces isomorphisms $E_1^{p0} F(X^B) \rightarrow E_1^{p0} F(X)$ for all p and all $X \in \mathcal{A}$.
- (5) $H^0 F$ is naturally isomorphic to $\psi\phi$.

Here, using the fact that $A//B$ -modules are A -modules, that is, using the adjoint j to ϕ , we construct a functor $\Gamma = E_1^{\bullet 0} F j$ from \mathcal{C} to cochain complexes in \mathcal{D} . This cannot be done in the G or S contexts. So we must include the cochain complex functor as part of the structure in our definition of a constructor. If F is a filtered cochain complex, we denote the component of total degree n by ${}^n F$, the p^{th} filtration by F^p , the submodule of filtration degree p and complementary degree q by F^{pq} and use similar notation for the terms of its spectral sequence $E(F)$. We always assume that ${}^n F^0 = {}^n F$ and that ${}^n F^{n+1} = 0$. The following definition generalises the one quoted above.

Definition 3.1. A spectral sequence constructor for the pair (ϕ, ψ) is a quadruple $(F, \Gamma, \eta, \gamma)$, where F is a functor from \mathcal{A} to filtered cochain complexes in \mathcal{D} , Γ is a functor from \mathcal{C} to cochain complexes in \mathcal{D} , η is a natural isomorphism $E_1^{\bullet 0}(F) \rightarrow \Gamma \phi$ and γ is a natural isomorphism $H^0(\Gamma) \rightarrow \psi$ such that

- (1) F is exact (in every filtration).
- (2) F is acyclic on injectives; that is, if Q is injective, then $H^n(FQ) = 0$ for $n > 0$ and for all p , $H^q(\bullet F^p Q / \bullet F^{p+1} Q) = 0$ for $q > 0$.
- (3) Γ is exact and acyclic on injectives.

From (1), it follows that $\bullet F^p / \bullet F^{p+r}$ is exact for all p, r . From γ being a natural isomorphism and (3), it follows that we have a unique family of natural isomorphisms $\gamma^p : H^p \Gamma \rightarrow R^p \psi$, the right derived functors of ψ , commuting with connecting homomorphisms, and with $\gamma^0 = \gamma$. We denote this family by γ . Furthermore,

$$H^0 F = E_{\infty}^{00} F = E_2^{00} F = H^0 E_1^{\bullet 0} F = H^0 \eta^{-1} \Gamma \phi,$$

and so

$$\gamma H^0(\eta) H^0 F = \gamma \phi = \theta;$$

that is, $\gamma H^0(\eta)$ is a natural isomorphism from $H^0 F$ to θ . It follows that $H^n F = R^n \theta$ for all n .

We now prove the results corresponding to Barnes [1, Theorem III.2.3, p. 42] and the lemmas leading up to that theorem, beginning with the analogue of [1, Theorem III.1.5, p. 38].

Lemma 3.2. *Let $(F, \Gamma, \eta, \gamma)$ be a spectral sequence constructor. There exists a unique family of natural transformations $\eta^{pq} : E_1^{pq}(F) \rightarrow \Gamma^p(R^q \phi)$ commuting with connecting homomorphisms and with $\eta^{p0} = \eta^p$. The η^{pq} are natural isomorphisms, satisfy $\eta^{p+1, q} d_1^{pq} = (-1)^q \delta \eta^{pq}$, where δ is the differential of Γ , and induce natural isomorphisms*

$$H^p(\eta^{\bullet q}) : E_2^{pq}(F) \rightarrow H^p(\Gamma(R^q \phi))$$

and

$$\gamma H^p(\eta^{\bullet q}) : E_2^{pq}(F) \rightarrow (R^p \psi)(R^q \phi).$$

Proof. Γ is exact by assumption. So $\{\Gamma^p(R^q \phi) | q = 0, 1, \dots\}$ is a connected sequence of functors, as is $\{E_1^{pq}(F) = H^q(F^p/F^{p+1}) | q = 0, 1, \dots\}$. Both vanish for $q > 0$ on injectives. By dimension-shifting, it follows that there exists a unique family of natural transformations $\eta^{pq} : E_1^{pq}(F) \rightarrow \Gamma^p(R^q \phi)$ extending the given transformation $\eta^p : E_1^{p0}(F) \rightarrow \Gamma^p \phi$. Since η^{p0} is a natural isomorphism, all the η^{pq} are natural isomorphisms. The argument of [1, pp. 38, 39] applies unchanged to give the result. \square

Definition 3.3. A natural transformation $\xi : (F, \Gamma, \eta, \gamma) \rightarrow (F', \Gamma', \eta', \gamma')$ of constructors is a pair $\xi_F : F \rightarrow F'$ and $\xi_\Gamma : \Gamma \rightarrow \Gamma'$ of natural transformations such that the diagrams

$$\begin{array}{ccc} E_1^{\bullet 0}(F) & \xrightarrow{\eta} & \Gamma\phi \quad \text{and} \quad H^0(\Gamma) \xrightarrow{\gamma} \psi \\ E_1^{\bullet 0}(\xi) \downarrow & & \downarrow \xi\phi \quad \quad \quad H^0(\xi) \downarrow \quad \quad \quad \downarrow \text{id} \\ E_1^{\bullet 0}(F') & \xrightarrow{\eta'} & \Gamma'\phi \quad \quad \quad H^0(\Gamma') \xrightarrow{\gamma'} \psi \end{array}$$

commute.

We shall omit the subscripts from ξ_F and ξ_Γ . Our next lemma is easier than [1, II.2.2] in that the transformation $\xi : \Gamma \rightarrow \Gamma'$ is given instead of having to be constructed.

Lemma 3.4. Let $\xi : (F, \Gamma, \eta, \gamma) \rightarrow (F', \Gamma', \eta', \gamma')$ be a natural transformation of constructors. Then the diagram

$$\begin{array}{ccc} E_1^{pq}(F) & \xrightarrow{\eta^{pq}} & \Gamma^p R^q \phi \\ \xi \downarrow & & \downarrow \xi R^q \phi \\ E_1^{pq}(F') & \xrightarrow{\eta'^{pq}} & \Gamma'^p R^q \phi \end{array}$$

commutes for all p and q .

Proof. $\xi R^\bullet \phi$ is a natural transformation of connected sequences of functors. Since in dimension $q = 0$, we have $\xi R^0 \phi = \eta'^{p0} \xi (\eta^{p0})^{-1}$, by dimension-shifting, we have $\xi R^q \phi = \eta'^{pq} \xi (\eta^{pq})^{-1}$ for all q . \square

Theorem 3.5. Let $\xi : (F, \Gamma, \eta, \gamma) \rightarrow (F', \Gamma', \eta', \gamma')$ be a natural transformation of constructors. Then ξ induces a natural isomorphism $\xi_E : E(F) \rightarrow E(F')$ of their spectral sequences, that is, $\xi_r^{pq} : E_r^{pq}(F) \rightarrow E_r^{pq}(F')$ is a natural isomorphism for all $r \geq 2$ and all p, q .

Proof. By assumption, $\gamma : H^0(\Gamma) \rightarrow \psi$ and $\gamma' : H^0(\Gamma') \rightarrow \psi$ are natural isomorphisms, and $\gamma' H^0(\xi) = \gamma$. Therefore $H^0(\xi) = (\gamma')^{-1} \gamma$ is a natural isomorphism. By dimension-shifting, it follows that $H^p(\xi) : H^p \Gamma \rightarrow H^p \Gamma'$ is a natural isomorphism for all p . The diagram

$$\begin{array}{ccc} E_1^{\bullet q}(F) & \xrightarrow{\eta} & \Gamma^\bullet R^q \phi \\ \xi_1^{\bullet q} \downarrow & & \downarrow \xi \\ E_1^{\bullet q}(F') & \xrightarrow{\eta'} & \Gamma'^\bullet R^q \phi \end{array}$$

is, up to sign, a commutative diagram of cochain complexes. (If q is odd, η and η' anticommute with the differentials.) Taking H^p of this, we get the commutative diagram

$$\begin{array}{ccc} E_2^{\bullet q}(F) & \xrightarrow{\eta} & (R^p \psi)(R^q \phi) \\ \xi_2^{\bullet q} \downarrow & & \downarrow H^p(\xi) \\ E_2^{\bullet q}(F') & \xrightarrow{\eta'} & (R^p \psi)(R^q \phi) \end{array}$$

in which η, η' and $H^p(\xi)$ are natural isomorphisms. It follows that ξ_2^{pq} is a natural isomorphism and so, that ξ_r^{pq} is a natural isomorphism for all $r \geq 2$. \square

4. GUICHARDET'S FIRST CONSTRUCTOR

For the A^e -module X , Guichardet defines the double complex

$$K^{pq}(X) = (\mathcal{F}(G^{p+1}, \text{Hom}(\otimes^{q+1} A^e, X))^B)^G$$

with appropriately defined differential, where $\mathcal{F}(U, V)$ denotes the set of functions from the set U to the set V . Following Guichardet [3], we set $I^n(X) = \text{Hom}(\otimes^{n+1} A^e, X)$ with differential

$$\begin{aligned} df(a_0 \otimes a'_0, \dots, a_{n+1} \otimes a'_{n+1}) \\ = a_0 f(a_1 \otimes a'_1, \dots, a_{n+1} \otimes a'_{n+1}) a'_0 \\ + \sum_{i=0}^n (-1)^{i+1} f(a_0 \otimes a'_0, \dots, a_i a_{i+1} \otimes a'_{i+1} a'_i, \dots, a_{n+1} \otimes a'_{n+1}), \end{aligned}$$

which gives a relatively injective resolution $I^\bullet(X)$ of X in \mathcal{A} . Also following Guichardet, we put $P_n = \otimes^{n+1} \mathfrak{K}G$ with action $g(g_0 \otimes \dots \otimes g_n) = gg_0 \otimes \dots \otimes gg_n$ and set $d(g_0 \otimes \dots \otimes g_n) = \sum_{i=0}^n (-1)^i (g_0 \otimes \dots \otimes \widehat{g_i} \otimes \dots \otimes g_n)$ and $\epsilon(g_0) = 1$. This makes P_\bullet a free resolution of \mathfrak{K} in \mathcal{C} . We then have

$$K^{pq}(X) = \text{Hom}_{\mathfrak{K}G}(P_p, I^q(X)^B).$$

Expressed in this way, it is the Grothendieck repeated (relatively) injective resolution construction for the spectral sequence of a composite functor discussed in Barnes [1, Chapter VII], with $\text{Hom}_{\mathfrak{K}}(P_\bullet, \)$ used as the relatively injective resolution functor on \mathcal{C} . For any relatively injective resolution functor I^\bullet and any projective resolution P_\bullet , setting $K^{pq}(X) = \text{Hom}_{\mathfrak{K}G}(P_p, I^q(X)^B)$ gives a constructor $(K, \Gamma, \eta, \gamma)$ with $\Gamma = \text{Hom}_{\mathfrak{K}G}(P_p, \)$, $\eta = \text{id}$ and $\gamma = \text{id}$. The spectral sequence constructed is independent (from the E_2 -level onward) of the choice of I^\bullet and of P_\bullet .

This constructor may also be regarded as an adaptation of the Cartan and Eilenberg pair of resolutions constructor discussed in Barnes [1, Chapter VI]. Since C -modules are not A^e -modules, we cannot use $\text{Hom}_{A^e}(P_p, I^q)$ as in the HS context, but use instead $\text{Hom}_C(P_p, (I^q)^B)$ which, in the HS context, is essentially the same.

Stefan in [5] establishes the conditions for the Grothendieck composite functor spectral sequence. To obtain a spectral sequence constructor, we have merely to assign functorially the resolutions used in the construction. If we assume that A is left and right B -projective or if we assume that \mathfrak{K} is a field, then we can use the I^n defined as above and any right C -module projective resolution P_\bullet of \mathfrak{K} .

5. GUICHARDET'S SECOND CONSTRUCTOR

For his second construction, Guichardet defines a filtration on the normalised standard complex ${}^\bullet N(A, X)$ where ${}^n N(A, X)$ is the subspace of $\text{Hom}_{\mathfrak{K}}(\otimes^n A, X)$ of functions f for which $f(a_1, \dots, a_n) = 0$ if any of the a_i is in $\mathfrak{K}1$, and

$$\begin{aligned} df(a_1, \dots, a_{n+1}) \\ = a_1 f(a_2, \dots, a_{n+1}) + \sum_{i=1}^n (-1)^i f(a_1, \dots, a_i a_{i+1}, \dots, a_{n+1}) \\ + (-1)^{n+1} f(a_1, \dots, a_n) a_{n+1}. \end{aligned}$$

The filtration on this complex is given by defining ${}^nN^0 = {}^nN$ and ${}^nN^p$ for $1 \leq p \leq n$ to be the subset of those functions f satisfying

$$f(a_1, \dots, a_q, g_1 b_1, \dots, g_p b_p) \\ = f(a_1, \dots, a_q, g_1, \dots, g_p) \alpha_{g_2 g_3 \dots g_p}^{-1}(b_1) \alpha_{g_3 \dots g_p}^{-1}(b_2) \dots \alpha_{g_p}^{-1}(b_{p-1}) b_p.$$

Guichardet takes for Γ the normalised standard complex and constructs a natural transformation from $E_1^{\bullet 0}(N)$ to $\Gamma\phi$ which, in [3, Lemme 3.11], he shows is a natural isomorphism. The conditions for a spectral sequence constructor are clearly satisfied. The purpose of this section is to generalise this to the S context.

To use the normalised standard complex in the S context, we must impose a further condition on the algebra A . The theory of the normalised standard complex (Cartan and Eilenberg [2, p. 176]) requires that the quotient $\bar{A} = A/\mathfrak{K}1$ be projective as \mathfrak{K} -module. We assume this in this section. Equivalently, we assume that there exists a \mathfrak{K} -linear map $\epsilon : A \rightarrow \mathfrak{K}$ such that $\epsilon(k1) = k$. This condition always holds if \mathfrak{K} is a field or if, as in the HS context, A is an augmented algebra.

An equivalent definition of ${}^nN^p$, also given by Guichardet, is meaningful in the S context. So we use it here but with sides reversed because of our use of right comodule algebras and right C -modules. For $p \geq 1$, we define ${}^nN^p$ to be the subset of ${}^nN(A, X)$ of those functions satisfying

$$(5.1) \quad f(ba_1, \dots, a_{n-1}, a_n) = bf(a_1, \dots, a_n)$$

and

$$(5.2) \quad f(a_1, \dots, a_{i-1}b, a_i, \dots, a_n) = f(a_1, \dots, a_{i-1}, ba_i, \dots, a_n) \quad \text{for } i = 2, \dots, p$$

for all $a_1, \dots, a_n \in A$ and $b \in B$.

The normalised standard complex ${}^\bullet N(C, Y)$ for a right C -module Y is that obtained by treating Y as a bimodule with left action $c \cdot y = \epsilon(c)y$. Thus, ${}^nN(C, Y)$ is the subspace of $\text{Hom}_{\mathfrak{K}}(\otimes^n C, Y)$ of functions f for which $f(c_1, \dots, c_n) = 0$ if any of the c_i is in $\mathfrak{K}1$, with the differential

$$df(c_1, \dots, c_{n+1}) = \epsilon(c_1)f(c_2, \dots, c_{n+1}) + \sum_{i=1}^n (-1)^i f(c_1, \dots, c_i c_{i+1}, \dots, c_{n+1}) \\ + (-1)^{n+1} f(c_1, \dots, c_n) c_{n+1}.$$

We put $T^{pq} = {}^pN(C, {}^qN(B, X))$ and write N^{pq} for ${}^{p+q}N^p$. Note that, although $N(B, X)$ is not, in general, a C -module, this does define \mathfrak{K} -modules T^{pq} .

For $f \in N^{pq}(A, X)$, we put

$$\Psi^{pq}(f)(c_1, \dots, c_p)(b_1, \dots, b_q) \\ = \sum r_{i_p}(c_p) \dots r_{i_1}(c_1) f(l_{i_1}(c_1), \dots, l_{i_p}(c_p), b_1, \dots, b_q).$$

That $\Psi^{pq} : N^{pq} \rightarrow T^{pq}$ is a well-defined \mathfrak{K} -linear map follows from the next lemma. We shorten the notation by writing \vec{a} for a string a_1, \dots, a_q of elements of A of any length. We further shorten notation by omitting unnecessary subscripts from the $r_i(c)$, $l_i(c)$.

Lemma 5.3. *If $f \in N^{pq}(A, X)$, then for $j = 1, \dots, p$, and all $\vec{a} \in A$, $b \in B$ and $c, c_1, \dots, c_p \in C$,*

- (1) $\sum r(c_j) \dots r(c_1) f(l(c_1), \dots, l(c_j), \vec{a})$ is independent of the choice of the $r(c_j)$ and $l(c_j)$.

$$\begin{aligned} (2) \quad & \sum br(c_j)r(c_{j-1}) \dots r(c_1)f(l(c_1), \dots, l(c_j), \vec{a}) \\ &= \sum r(c_j) \dots r(c_1)f(l(c_1), \dots, l(c_j)b, \vec{a}). \\ (3) \quad & \sum r(c)r(c_j) \dots r(c_1)f(l(c_1), \dots, l(c_j)l(c), \vec{a}) \\ &= \sum r(c_jc) \dots r(c_1)f(l(c_1), \dots, l(c_jc), \vec{a}). \end{aligned}$$

Proof. For $a, a' \in A$, we put

$$g_j(a \otimes_B a') = \sum ar(c_{j-1}) \dots r(c_1)f(l(c_1), \dots, l(c_{j-1}), a', \vec{a}).$$

By the condition (5.1), g_1 is well defined. Thus (1) holds for $j = 1$. Also, by putting

$$a \otimes_B a' = \sum br(c_1) \otimes_B l(c_1) = \sum r(c_1) \otimes_B l(c_1)b,$$

by Lemma 2.1(a), we see that (2) holds for $j = 1$. We use induction over j .

For $1 < j \leq p$, we have

$$\begin{aligned} & \sum abr(c_{j-1}) \dots r(c_1)f(l(c_1), \dots, l(c_{j-1}), a', \vec{a}) \\ &= \sum ar(c_{j-1}) \dots r(c_1)f(l(c_1), \dots, l(c_{j-1})b, a', \vec{a}) \\ &= \sum ar(c_{j-1}) \dots r(c_1)f(l(c_1), \dots, l(c_{j-1}), ba', \vec{a}) \end{aligned}$$

by the induction hypothesis that (2) holds for $j - 1$ and condition (5.2). Thus g_j is well defined. Putting $a \otimes_B a' = \sum r(c_j) \otimes_B l(c_j)$ gives the assertion (1). Putting $a \otimes_B a' = \sum br(c_j) \otimes_B l(c_j)$ and using Lemma 2.1(a) gives (2). Putting $a \otimes_B a' = \sum r(c)r(c_j) \otimes_B l(c_j)l(c)$ and using Lemma 2.1(f) gives (3). \square

Lemma 5.4. Ψ defines a natural cochain map $\Psi_0^{p\bullet} : E_0^{p\bullet} \rightarrow T^{p\bullet}$.

Proof. If $f \in {}^nN^{p+1}$, then

$$\begin{aligned} \Psi^{pq}(f)(c_1, \dots, c_p)(b_1, \dots, b_q) &= \sum r(c_p) \dots r(c_1)f(l(c_1), \dots, l(c_p)b_1, 1, b_2, \dots, b_q) \\ &= 0. \end{aligned}$$

Since ${}^nE_0^p = {}^nN^p / {}^nN^{p+1}$, Ψ^{pq} defines a \mathfrak{K} -linear map $\Psi_0^{pq} : E_0^{pq} \rightarrow T^{pq}$. Consider the expression for $\sum (d_A f)(l(c_1), \dots, l(c_p), b_1, \dots, b_{q+1})$. For those terms in which the string of $l(c_i)$'s is reduced in length, we get b_1 in the p^{th} place; so those terms are 0. Thus,

$$\begin{aligned} & (\Psi^{p,q+1}d_A f)(c_1, \dots, c_p)(b_1, \dots, b_{q+1}) \\ &= (-1)^p \sum r(c_p) \dots r(c_1)f(l(c_1), \dots, l(c_p)b_1, \dots, b_{q+1}) \\ &\quad + \sum (-1)^{p+i} r(c_p) \dots f(\dots, (b_i b_{i+1}), \dots) \\ &\quad + (-1)^{p+q+1} \sum r(c_p) \dots f(l(c_1), \dots, b_q)b_{q+1} \\ &= (-1)^p d_B((\Psi^{pq} f)(c_1, \dots, c_p))(b_1, \dots, b_{q+1}) \end{aligned}$$

by applying Lemma 5.3(2) to the first term. The result follows, the naturality being obvious. \square

We are trying to construct a spectral sequence constructor using $N_A = N(A, \quad)$ with the Guichardet filtration as the filtered complex functor. Clearly, we can set $\Gamma = N(C, \quad)$ and $\gamma = \text{id} : H^0(C, \quad) \rightarrow \psi$. We still need a natural isomorphism $\eta : E_1^{\bullet 0}(N_A) \rightarrow \Gamma\phi$. Applying H^q to the natural cochain map $\Psi_0^{p\bullet}$ gives a natural map $\eta^{pq} : E_1^{pq} \rightarrow {}^pN(C, H^q(B, \quad)) = \Gamma^p(H^q(B, \quad))$. We must first show that $\eta^{\bullet 0}$ is an isomorphism of cochain complexes.

Lemma 5.5. $\eta^{\bullet 0} : E_1^{\bullet 0}(X) \rightarrow \Gamma^{\bullet} X^B$ is a map of cochain complexes.

Proof. An element of $E_1^{p0}(X)$ is represented by a function $f \in {}^pN^p(A, X)$ such that $df \in {}^{p+1}N^{p+1}$. For $f \in {}^pN^p$, every term t in df satisfies (5.1) and (5.2) for all i except the term $t(a_1, \dots, a_{p+1}) = f(a_1, \dots, a_p)a_{p+1}$ for which (5.2) may fail for $i = p + 1$. Thus the requirement that $df \in {}^{p+1}N^{p+1}$ imposes the one extra condition that $f(a_1, \dots, a_p b)a_{p+1} = f(a_1, \dots, a_p)ba_{p+1}$, that is, $f(a_1, \dots, a_p b) = f(a_1, \dots, a_p)b$. For such an f , we have, writing \vec{c} for c_1, \dots, c_{p+1} ,

$$\begin{aligned} (\Psi^{p+1,0} d_A f)(\vec{c}) &= \sum r(c_{p+1}) \dots r(c_1) df(l(c_1), \dots, l(c_{p+1})) \\ &= \sum r(c_{p+1}) \dots r(c_1) l(c_1) f(l(c_2), \dots, l(c_{p+1})) \\ &\quad + \sum (-1)^i r(c_{p+1}) \dots r(c_1) f(l(c_1), \dots, (c_i)l(c_{i+1}), \dots, l(c_{p+1})) \\ &\quad + \sum (-1)^{p+1} r(c_{p+1}) \dots r(c_1) f(l(c_1), \dots, l(c_p))l(c_{p+1}). \end{aligned}$$

By Lemma 2.1(c),

$$\begin{aligned} \sum r(c_{p+1}) \dots r(c_1) l(c_1) f(l(c_2), \dots, l(c_{p+1})) \\ = \epsilon(c_1) \sum r(c_{p+1}) \dots r(c_2) f(l(c_2), \dots, l(c_{p+1})). \end{aligned}$$

By Lemma 5.3(3),

$$\begin{aligned} \sum (-1)^i r(c_{p+1}) \dots r(c_1) f(l(c_1), \dots, (c_i)l(c_{i+1}), \dots, l(c_{p+1})) \\ = \sum (-1)^i r(c_{p+1}) \dots r(c_i c_{i+1}) \dots r(c_1) f(l(c_1), \dots, (c_i c_{i+1}), \dots, l(c_{p+1})). \end{aligned}$$

Also,

$$\begin{aligned} \sum r(c_{p+1}) \dots r(c_1) f(l(c_1), \dots, l(c_p))l(c_{p+1}) \\ = \left(\sum r(c_p) \dots r(c_1) f(l(c_1), \dots, l(c_p)) \right) \cdot c_{p+1}. \end{aligned}$$

Thus $\Psi^{p+1,0} d_A f = d_C \Psi^{p0} f$ and the result follows. \square

For $g \in {}^pN(C, X^B)$, we define $\Phi g \in {}^pN(A, X)$ by

$$(\Phi g)(a_1, \dots, a_p) = \sum a_1^0 \dots a_p^0 g(a_1^1, \dots, a_p^1);$$

writing the comodule structure indices as superscripts, $\Delta a_i = \sum a_i^0 \otimes a_i^1$.

Lemma 5.6. $\Phi g \in {}^pN^p(A, X)$ and $d_A(\Phi g) \in {}^{p+1}N^{p+1}(A, X)$.

Proof. For $b \in B$, we have

$$\Phi g(ba_1, \dots, a_p) = \sum ba_1^0 \dots a_p^0 g(a_1^1, \dots, a_p^1)$$

since $\Delta_A(ba_1) = \sum ba_1^0 \otimes a_1^1$. Thus condition (5.1) is satisfied. Also,

$$\begin{aligned} (\Phi g)(a_1, \dots, a_i b, a_{i+1}, \dots, a_p) \\ = \sum a_1^0 \dots a_i^0 ba_{i+1}^0 \dots a_p^0 g(a_1^1, \dots, a_p^1) \\ = (\Phi g)(a_1, \dots, a_i, ba_{i+1}, \dots, a_p). \end{aligned}$$

Thus (5.2) is satisfied for all i and $\Phi g \in {}^pN^p(A, X)$. Since $g(a_1^1, \dots, a_p^1) \in X^B$,

$$\begin{aligned} (\Phi g)(a_1, \dots, a_p b) &= \sum a_1^0 \dots a_p^0 b g(a_1^1, \dots, a_p^1) \\ &= \sum a_1^0 \dots a_p^0 g(a_1^1, \dots, a_p^1) b = \Phi g(a_1, \dots, a_p) b \end{aligned}$$

and it follows that $d_A \Phi g \in {}^{p+1}N^{p+1}(A, X)$. □

Lemma 5.7. *For $g \in {}^pN(C, X^B)$, $\Psi^{p0} \Phi g = g$.*

Proof. For any \mathfrak{K} -linear function $t : C \rightarrow X$, setting $u(a \otimes_B a' \otimes c) = aa't(c)$ for $a, a' \in A$ and $c \in C$ defines a \mathfrak{K} -linear function $u : A \otimes_B A \otimes C \rightarrow X$. By Lemma 2.1(d),

$$\begin{aligned} \sum r(c) l(c)^0 t(l(c)^1) &= u\left(\sum r(c) \otimes_B l(c)^0 \otimes l(c)^1\right) \\ &= u\left(\sum r(c^1) \otimes_B l(c^1) \otimes c^2\right) \\ &= \sum r(c^1) l(c^1) t(c^2) \\ &= \sum \epsilon(c^1) t(c^2) \text{ by Lemma 2.1(c)} \\ &= t(c). \end{aligned}$$

Using this with $t(c) = \sum l(c_2)^0 \dots l(c_p)^0 g(c, l(c_2)^1, \dots, l(c_p)^1)$, we have

$$\begin{aligned} (\Psi^{p0} \Phi g)(c_1, \dots, c_p) &= \sum r(c_p) \dots r(c_1) (\Phi g)(l(c_1), \dots, l(c_p)) \\ &= \sum r(c_p) \dots r(c_1) l(c_1)^0 \dots l(c_p)^0 g(l(c_1)^1, \dots, l(c_p)^1) \\ &= \sum r(c_p) \dots r(c_2) l(c_2)^0 \dots l(c_p)^0 g(c_1, l(c_2)^1, \dots, l(c_p)^1). \end{aligned}$$

Repeating this argument gives the result. □

Lemma 5.8. *If $f \in {}^pN^p(A, X)$ and $d_A f \in {}^{p+1}N^{p+1}(A, X)$, then $\Phi \Psi^{p0} f = f$.*

Proof. Setting $u(a \otimes_B a') = \sum ar(a_{p-1}^1) \dots r(a_1^1) f(l(a_1^1), \dots, l(a_{p-1}^1), a')$ defines a \mathfrak{K} -linear function $u : A \otimes_B A \rightarrow X$ by Lemma 5.3(b) and condition 5.2. We have

$$\begin{aligned} (\Phi \Psi^{p0} f)(a_1, \dots, a_p) &= \sum a_1^0 \dots a_p^0 (\Psi^{p0} f)(a_1^1, \dots, a_p^1) \\ &= \sum a_1^0 \dots a_p^0 r(a_p^1) \dots r(a_1^1) f(l(a_1^1), \dots, l(a_p^1)) \\ &= \sum a_1^0 \dots a_{p-1}^0 u(a_p^0 r(a_p^1) \otimes_B l(a_p^1)) \\ &= \sum a_1^0 \dots a_{p-1}^0 u(1 \otimes_B a_p) \text{ by Lemma 2.1(b)} \\ &= \sum a_1^0 \dots a_{p-1}^0 1r(a_{p-1}^1) \dots r(a_1^1) f(l(a_1^1), \dots, l(a_{p-1}^1), a_p). \end{aligned}$$

Repeating this argument gives the result. □

Corollary 5.9. *The η^{p0} are isomorphisms.*

Proof. E_1^{p0} is the set of $f \in {}^pN(A, X)$ with $d_A f \in {}^{p+1}N^{p+1}(A, x)$, and η^{p0} is the restriction $\Psi^{p0} | E_1^{p0} \rightarrow {}^pN(C, X^B)$. By Lemma 5.7, it is surjective and, by Lemma 5.8, it is injective. □

Theorem 5.10. *Suppose $\bar{A} = A/\mathfrak{K}1$ is projective as \mathfrak{K} -module and that A/B is projective as left B -module. Then $(N_A, \Gamma, \eta, \gamma)$ is a spectral sequence constructor for (ϕ, ψ) .*

Proof. We have to show that the conditions (1), (2), (3) of Definition 3.1 are satisfied. Since

$${}^nN_A(X) = \text{Hom}(\otimes^n(\bar{A}), X) = \text{Hom}(\bar{A}, \text{Hom}(\otimes^{n-1}\bar{A}, X))$$

and \bar{A} is \mathfrak{K} -projective, nN is an exact functor. A function $f \in {}^nN^1$ satisfies the further condition $f(ba_1, \dots, a_n) = bf(a_1, \dots, a_n)$ for all $b \in B$. In particular, $f(a_1, \dots, a_n) = 0$ if $a_1 \in B$. Thus

$${}^nN_A^1(X) = \text{Hom}(\otimes^{n-1}\bar{A}, \text{Hom}_B(A/B, X)).$$

Thus ${}^nN_A^1$ is an exact functor. Similarly,

$$\begin{aligned} {}^nN_A^p(X) &= \text{Hom}(\otimes^q \bar{A}, \text{Hom}_B(\otimes_B^p A/B, X)) \\ &= \text{Hom}(\otimes^q \bar{A}, \text{Hom}_B(A/B, \text{Hom}_B(\otimes_B^{p-1} A/B, X))) \end{aligned}$$

and by induction over p , ${}^nN_A^p$ is exact. Thus condition (1) holds.

Let Q be an injective A^e -module. Then $H^n(N_A Q) = 0$ for $n > 0$ by the usual theory of the normalised standard complex. We have to show that

$$H^q(\bullet N_A^p Q / \bullet N_A^{p+1} Q) = 0$$

for $q > 0$. But

$$H^q(\bullet N_A^p Q / \bullet N_A^{p+1} Q) = E_1^{pq} N_A(Q) \simeq {}^pN(C, H^q(B, Q))$$

by Corollary 5.9. But $H^q(B, Q) = 0$ for $q > 0$ since Q is injective as B^e -module. Thus condition (2) holds.

Since $\Gamma(Y) = N(C, Y)$, condition (3) holds. \square

In the discussion of the filtered normalised complex in Barnes [1, Chapter IV], the corresponding extra assumption that A/B be projective as right B -module was needed. If in the HS context, A is a Hopf algebra, then it can be regarded as a $A//B$ -comodule algebra. The Guichardet filtration is not the same as that given by Hochschild and Serre, but by the result of the next section, the two filtrations give the same spectral sequence.

6. UNIQUENESS OF THE SPECTRAL SEQUENCE

As in Barnes [1, Chapter X], we construct for each cardinal α , a cofree functor which, restricted to the subcategory \mathcal{A}_α of objects of cardinality less than α , is a spectral sequence constructor. (This use of the subcategory \mathcal{A}_α is necessary because a cofree functor with injective model M and injective basis (M, U) only has the desired properties with respect to modules embeddable in M .) From the existence of this cofree functor, we deduce as in [1, Chapter X], that all spectral sequence constructors construct the same spectral sequence. We need one technical lemma to get around the difficulty caused by C -modules not being A^e -modules. For this, we again need the assumption that \bar{A} is \mathfrak{K} -projective, that is, that there exists a \mathfrak{K} -module homomorphism $\epsilon : A \rightarrow \mathfrak{K}$ with $\epsilon(1) = 1$.

Lemma 6.1. *For every C -module Y , there exists an injective A^e -module Q such that Y can be embedded in Q^B .*

Proof. We first make $Z = \text{Hom}(A, Y)$ a C -module by defining $(f \cdot c)(a) = f(a)c$ for $c \in C$, $a \in A$ and $f \in \text{Hom}(A, Y)$. We construct an embedding $i : Y \rightarrow Z$ by setting $(iy)(a) = \epsilon(a)y$ for $y \in Y$ and $a \in A$. So defined, i is a C -module homomorphism, because

$$(i(yc))(a) = \epsilon(a)yc = (\epsilon(a)y)c = (iy)(a)c = ((iy) \cdot c)(a)$$

for $g \in G$, $a \in A$ and $f \in \text{Hom}(A, Y)$. It is clearly injective.

Next, we use the standard embedding of Z in the coinduced C -module $W = \text{Hom}_{\mathfrak{K}}(C, Z)$, defining $\sigma : Z \rightarrow W$ by setting $\sigma(z)(c) = zc$ for $z \in Z$ and $c \in C$. We now have an embedding of Y in $\text{Hom}(C, \text{Hom}(A, Y))$. By Lemma 2.4, we have an embedding of Y in X^B where X is the coinduced A^e -module $\text{Hom}(A^e, Y)$. Taking any embedding of X in an injective A^e -module Q , we get an embedding of Y in Q^B . \square

We use the theory of cofree functors developed in [1, Chapter X]. Our spectral sequence constructors consist of two functors and two natural transformations instead of the single functor used in the HS context. To accommodate this, we shall say that the pair (F, Γ) of functors, F defined on \mathcal{A} and Γ defined on \mathcal{C} , is simple cofree on the basis (M, U, V) if F is cofree on (M, U) and Γ is cofree on $(\phi(M), V)$.

Theorem 6.2. *Let C be a Hopf algebra over \mathfrak{K} and let A be a right C -comodule algebra with $B = A^{\text{co}C}$. Suppose A, B and C are \mathfrak{K} -projective and that A/B is C -Galois. Let \mathcal{A} be the category of left A^e -modules, \mathcal{C} the category of right C -modules and let \mathcal{D} be the category of \mathfrak{K} -modules. Let $\phi : \mathcal{A} \rightarrow \mathcal{C}$ and $\psi : \mathcal{C} \rightarrow \mathcal{D}$ be the functors defined by $\phi(X) = X^B$ and $\psi(Y) = Y^C$ for $X \in \mathcal{A}$ and $Y \in \mathcal{C}$. Suppose $A/\mathfrak{K}1$ is \mathfrak{K} -projective and that A is both left and right B -flat. Then, for any cardinal α , there exists a simple cofree pair (T, Γ) with injective model M and injective basis (M, U, V) , and natural transformations η, γ such that on the subcategory \mathcal{A}_α of objects of \mathcal{A} of cardinality less than α , $(T, \Gamma, \eta, \gamma)$ is a spectral sequence constructor for (ϕ, ψ) .*

Proof. By replacing α by a suitable larger limit cardinal, we may suppose that every object in \mathcal{A}_α has an injective resolution in \mathcal{A}_α , and that every object of \mathcal{C}_α likewise has an injective resolution in \mathcal{C}_α . There exists an injective module X in \mathcal{A} such that every object of \mathcal{A}_α can be embedded in X . Likewise, there exists an injective module Y in \mathcal{C} such that every object of \mathcal{C}_α can be embedded in Y . By Lemma 6.1, there exists an injective module Q in \mathcal{A} such that Y can be embedded in $\phi(Q)$. Putting $M = X \oplus Q$, we obtain an injective module M such that every module in \mathcal{A}_α has an injective resolution, all of whose terms can be embedded in M , and every module in \mathcal{C}_α has an injective resolution all of whose terms can be embedded in $\phi(M)$. By [1, Lemma X.3.2, p. 101], there exists a simple cofree functor Γ from \mathcal{C} to cochain complexes in \mathcal{D} with basis $(\phi(M), V)$ for some injective V , and natural transformation $\gamma : H^\bullet(\Gamma) \rightarrow R^\bullet\psi$ which, on \mathcal{C}_α , is a natural isomorphism. The construction of T and the proof of the result now follows exactly as for [1, Theorem X.5.3, p. 107]. \square

Theorem 6.3. *Let C be a Hopf algebra over \mathfrak{K} and let A be a right C -comodule algebra with $B = A^{\text{co}C}$. Suppose A, B and C are \mathfrak{K} -projective and that A/B is C -Galois. Let \mathcal{A} be the category of left A^e -modules, \mathcal{C} the category of right C -modules and let \mathcal{D} be the category of \mathfrak{K} -modules. Let $\phi : \mathcal{A} \rightarrow \mathcal{C}$ and $\psi : \mathcal{C} \rightarrow \mathcal{D}$ be the functors defined by $\phi(X) = X^B$ and $\psi(Y) = Y^C$ for $X \in \mathcal{A}$ and $Y \in \mathcal{C}$.*

Suppose $A/\mathcal{K}1$ is \mathcal{K} -projective and that A is both left and right B -flat. Suppose $F = (F, \Gamma_F, \eta_F, \gamma_F)$ and $F' = (F', \Gamma'_{F'}, \eta'_{F'}, \gamma'_{F'})$ are spectral sequence constructors for (ϕ, ψ) . Then F and F' construct canonically isomorphic spectral sequences from the E_2 -level onward.

Proof. The argument of [1, Theorem X.5.4, p. 109] applies unchanged. \square

REFERENCES

- [1] D. W. Barnes, *Spectral sequence constructors in algebra and topology*, Mem. Amer. Math. Soc. **53** (1985). MR **86e**:55032
- [2] H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956. MR **17**:1040e
- [3] A. Guichardet, *Suites spectrales à la Hochschild-Serre pour les produits croisés d'algèbres et de groupes*, J. Algebra **235** (2001), 744–765. MR **2001m**:16013
- [4] H. J. Schneider, *Representation theory of Hopf Galois extensions*, Hopf algebras, Israel J. Math. **72** (1990), 196–231. MR **92d**:16047
- [5] D. Stefan, *Hochschild cohomology on Hopf Galois extensions*, J. Pure Appl. Algebra **103** (1995), 221–233. MR **96h**:16013

1 LITTLE WONGA ROAD, CREMORNE NSW 2090, AUSTRALIA

E-mail address: donb@netspace.net.au

FORMALITY IN AN EQUIVARIANT SETTING

STEVEN LILLYWHITE

ABSTRACT. We define and discuss G -formality for certain spaces endowed with an action by a compact Lie group. This concept is essentially formality of the Borel construction of the space in a category of commutative differential graded algebras over $R = H^\bullet(BG)$. These results may be applied in computing the equivariant cohomology of their loop spaces.

1. INTRODUCTION

In this paper we consider G -spaces and give formality results for them in an equivariant category. More specifically, given a G -space M , we discuss formality of the Borel construction $EG \times_G M$ or, equivalently, formality of the complex $A_G^\bullet(M)$ of equivariant differential forms. However, in the equivariant setting, the map $M \rightarrow \{pt.\}$ is replaced by $EG \times_G M \rightarrow BG$, and consequently all the commutative differential graded algebras involved are naturally R -algebras, where $R = H^\bullet(BG)$. Thus formality may be considered in the category of commutative differential graded R -algebras. We shall also consider the augmented case, corresponding to equivariant base points, which are the same thing as fixed points of the group action. We should like to call a G -space M “equivariantly formal” when its Borel construction is formal in the above sense. However, the term “equivariant formality” has come to be used to describe the degeneration of the spectral sequence of the fibration $M \rightarrow EG \times_G M \rightarrow BG$, owing to the pervasive influence of [11]. Thus we shall adopt the terminology “ G -formal” in this paper.

We give some general results concerning G -formality of products and wedges and reductions to subgroups. This is followed by several examples of G -formal spaces, including compact Kähler manifolds and formal elliptic spaces, among others. Of course, we must make appropriate assumptions on the G -actions of these spaces for the results to hold.

As an application of these results, we compute the equivariant cohomology of loop spaces. (If M is a G -space, then so is the loop space of M in the obvious way.) Our motivation comes from considering the cohomology of symplectic quotients of loop spaces, see [18], although the results are of general topological interest. We shall use an “equivariant” bar complex to compute the equivariant cohomology of the loop space. If the G -space M is G -formal, then the bar complex, which is generally a double complex, loses a differential and becomes a single complex, allowing for some easier calculations. In the last section we compute an example.

Received by the editors January 1, 2002.

2000 *Mathematics Subject Classification.* Primary 55P62; Secondary 55N91, 18G55, 57T30.

Key words and phrases. Rational homotopy theory, equivariant cohomology, bar complexes, loop spaces, homotopical algebra.

In an appendix, we discuss bar complexes and Eilenberg-Moore theory concerning the pull-back of a fibration. We also consider equivariant versions of these results, which are used in several of the proofs in the main body of the paper.

In what follows, we shall generally assume that G is a compact, connected Lie group and that all spaces are connected. Whenever we need to use the localization theorem in equivariant cohomology, we shall assume that the spaces under consideration are of the homotopy type of finite-dimensional G -CW-complexes, and furthermore that they have finitely many connective orbit types, meaning that the set $\{[G_x^0] \mid x \in M\}$ is finite, where G_x is the stabilizer subgroup at x , G_x^0 is the connected component of the identity, and $[G_x^0]$ denotes the set of conjugacy classes in G . This latter condition is automatically satisfied, by the way, if M is compact or if $G = S^1$.

I would like to extend my appreciation to Chris Allday, who took the time to read the manuscript and offered advice on several key points. In particular, Proposition 4.7 is due to him.

2. $kCDGA$ AND FORMALITY

In this section we recall some important facts about the category of commutative differential graded algebras, the notion of formality, and the connection with rational homotopy theory. We shall assume for now that our algebras are k -algebras, where k is a field of characteristic zero. We shall denote by $kCDGA^\circ$ the category of commutative differential graded k -algebras that are concentrated in non-negative degrees and have a differential that raises the degree by one. We assume further that $H^0(A) \approx k$, for all A in $kCDGA^\circ$. We shall denote by $kCDGA$ the category of algebras in $kCDGA^\circ$ that are augmented over k (i.e., there exists for each A a map $\varepsilon : A \rightarrow k$, with k concentrated in degree zero), together with augmentation-preserving maps for morphisms. We shall call an object of $kCDGA$ (resp. $kCDGA^\circ$) a $kCDGA$ (resp. $kCDGA^\circ$).

We recall Quillen's abstract approach to homotopy theory, [22], [23]. He begins by defining the notion of a closed model category. A closed model category is a category, \mathcal{C} , with 3 distinguished classes of morphisms, called cofibrations, fibrations, and weak equivalences, which satisfy a number of axioms. The homotopy category, $Ho \mathcal{C}$, is defined to be the localization of \mathcal{C} with respect to the class of weak equivalences. Quillen introduces a notion of homotopy and shows that $Ho \mathcal{C}$ is equivalent to the more concrete category $ho \mathcal{C}$ which has for its objects the cofibrant/fibrant objects of \mathcal{C} , and for its morphisms the homotopy classes of maps. We point out the important fact that two objects X and Y in $Ho \mathcal{C}$ are isomorphic if and only if there exists a chain (in \mathcal{C}) of weak equivalences

$$(1) \quad X \leftarrow Z_1 \rightarrow Z_2 \leftarrow \cdots \leftarrow Z_n \rightarrow Y.$$

In [4] it is shown that the categories $kCDGA^\circ$ and $kCDGA$ are closed model categories where the weak equivalences are the quasi-isomorphisms (maps that induce an isomorphism on cohomology), fibrations are the surjective morphisms, and cofibrations are maps that satisfy the following lifting condition: a map f is a

cofibration if for every commutative diagram

$$\begin{array}{ccc} X & \longrightarrow & V \\ f \downarrow & & p \downarrow \\ Y & \longrightarrow & W \end{array}$$

with p a fibration and weak equivalence, there is a map from Y to V making the diagram commute. (Actually, in [4], the authors do not assume that $H^0(A) \approx k$ for all algebras A . We have included this assumption for ease of presentation, but the difference is slight.)

Given a closed model category, \mathcal{C} , with initial object $*$, an object B is called *cofibrant* if the map $* \rightarrow B$ is a cofibration. B is called a *cofibrant model* for A if B is cofibrant and there exists a weak equivalence $B \rightarrow A$. It follows from the axioms for a closed model category that every object in a closed model category has a cofibrant model. Moreover, there are various lifting and homotopy results associated with cofibrant algebras; see [4], section 6. We mention one here. If $\varphi : B_1 \rightarrow B_2$ is a quasi-isomorphism, and we have a map $f : A \rightarrow B_2$ with A cofibrant, then there exists a lift $\tilde{f} : A \rightarrow B_1$ such that $\varphi \tilde{f} \simeq f$, where \simeq denotes homotopy.

Note that $kCDGA$ is pointed with point object k . The homotopy groups of a $kCDGA$ A are defined to be

$$\pi^n A \stackrel{\text{def}}{=} H^n(\bar{A}/(\bar{A})^2),$$

where $\bar{A} = \ker \varepsilon$, for $\varepsilon : A \rightarrow k$ a given augmentation of A .

If $f : B_1 \rightarrow B_2$ is a weak equivalence of cofibrant $kCDGA$'s, then $f_* : \pi^\bullet B_1 \rightarrow \pi^\bullet B_2$ is an isomorphism. Thus, if we define $\Pi^n(A) \stackrel{\text{def}}{=} \pi^n(B)$ for B a cofibrant model of A , then $\Pi^n(A)$ is well-defined up to isomorphism. Moreover, if $f : A_1 \rightarrow A_2$ is a map of $kCDGA$'s, then f induces a unique homotopy class of maps $f : B_1 \rightarrow B_2$, for fixed choices of cofibrant models B_1, B_2 of A_1, A_2 , respectively. It follows that there is a unique map $f_* : \Pi(A_1) \rightarrow \Pi(A_2)$. Thus Π is functorial, and different choices of cofibrant models yield naturally isomorphic such functors.

In $kCDGA$, there is a special class of cofibrant models called minimal models. A minimal model of an algebra A is defined to be a cofibrant model, $\mathcal{M} \rightarrow A$, that is connected ($\mathcal{M}^0 \approx k$), and such that the induced differential on $\bar{\mathcal{M}}/(\bar{\mathcal{M}})^2$ is zero. It can be shown that each algebra in $kCDGA$ has a minimal model, unique up to isomorphism. If M is a path-connected topological space, the (pseudo-dual) k -homotopy groups of M are defined to be $\Pi^n(M) \stackrel{\text{def}}{=} \Pi^n(A^\bullet(M)) = \pi^n(\mathcal{M})$, where \mathcal{M} is a minimal model for $A^\bullet(M)$. Here, $A^\bullet(M)$ denotes the Sullivan-de Rham complex, which is a $\mathbb{Q}CDGA$; see, for example, [3] for the definition. If M is a smooth manifold, we may also use the ordinary de Rham complex, taking k to be \mathbb{R} .

Halperin has explicitly identified the cofibrations (and hence cofibrant objects) in $kCDGA$. Cofibrations are the so-called KS-extensions, and the cofibrant objects are the KS-complexes. Since these notions will be important to us, we give their definitions here; see [14] or [3].

Definition 2.1. A map $f : A \rightarrow B$ of $kCDGA$'s is said to be a *KS-extension* if there exists a well-ordered subset $E \subset B$, $E = \{x_\alpha\}$, such that $A \otimes \bigwedge(E) \rightarrow B$ is an isomorphism of commutative graded algebras, where $\bigwedge(E)$ denotes the free

graded commutative algebra on the set E , and the map is induced by f and the inclusion of $E \subset B$. Identifying B with $A \otimes \bigwedge(E)$, the differential on B satisfies

- (1) $d_B(a \otimes 1) = d_A(a) \otimes 1$,
- (2) $d_B(1 \otimes x_\alpha) \in A \otimes \bigwedge(E_{<\alpha})$,

where $E_{<\alpha} = \{x_\beta \mid \beta < \alpha\}$. If E also satisfies $\deg(x_\alpha) > 0 \forall x_\alpha \in E$, and $\deg(x_\beta) < \deg(x_\alpha) \Rightarrow \beta < \alpha$, then f is called a *minimal KS-extension*. If $A = k$, then we replace the word “extension” by the word “complex” in the definition, obtaining the notion of *KS-complex*. (A minimal KS-complex is the same thing as a minimal algebra, defined above.) A minimal KS-extension in which A is also minimal is called a Λ -*minimal Λ -extension*. Note that in a (minimal) KS-extension, $\bigwedge(E)$ is a (minimal) KS-complex, with differential such that $\varepsilon \otimes 1 : A \otimes \bigwedge(E) \rightarrow \bigwedge(E)$ is a map of $k\text{CDGA}^o$'s, where ε is the augmentation of A . Moreover, all of these maps may be made compatible with augmentations.

If A is a $k\text{CDGA}$, then its cohomology, $H(A)$, may be considered to be a $k\text{CDGA}$ with zero differential.

Definition 2.2. A is said to be *formal* if $A \approx H(A)$ in $Ho(k\text{CDGA})$.

It is easy to see that this definition is equivalent to the following two.

Lemma 2.3. Consider the category $k\text{CDGA}$. The following are equivalent:

- (1) A is formal.
- (2) There is a diagram

$$A \leftarrow B \rightarrow H(A),$$

where the maps are weak equivalences and B is a cofibrant model for A . (In particular, we may pick B to be minimal.)

- (3) There is a chain of quasi-isomorphisms

$$A \leftarrow A_1 \rightarrow A_2 \leftarrow \cdots \leftarrow A_n \rightarrow H(A).$$

This theory has an important application to rational homotopy theory. It turns out that the homotopy category of rational finite \mathbb{Q} -type nilpotent spaces is equivalent to the homotopy category of the full subcategory of $\mathbb{Q}\text{CDGA}$ consisting of algebras A with ΠA of finite type, [4]. Thus we may “do” rational homotopy theory in a category of differential graded algebras. As an example, if X is a path-connected, simply-connected topological space of finite \mathbb{Q} -type, then there is a natural isomorphism

$$\Pi^n(A^\bullet(X)) \approx \text{Hom}_{\mathbb{Q}}(\pi_n(X) \otimes \mathbb{Q}, \mathbb{Q}),$$

where $A^\bullet(X)$ is the $\mathbb{Q}\text{CDGA}$ of Sullivan-de Rham differential forms on X . If X is a smooth manifold, the same statement for homotopy groups holds if we use instead the de Rham algebra $A^\bullet(X)$ and replace \mathbb{Q} coefficients by \mathbb{R} , or \mathbb{C} . There is not a corresponding equivalence of homotopy categories over \mathbb{R} or \mathbb{C} , however.

A path-connected topological space is said to be *formal* if its Sullivan-de Rham algebra $A^\bullet(X)$ is formal. If X is a smooth manifold, we may use the de Rham algebra and real or complex coefficients. However, a well-known result in rational homotopy theory states that formality over \mathbb{R} or \mathbb{C} implies formality over \mathbb{Q} ; see, for example, [15].

Formal spaces include compact Kähler manifolds and many homogeneous spaces, including compact globally symmetric spaces. Products, wedges, and connected sums of formal spaces are again formal. The topological consequences of formality include the vanishing of all Massey products. Moreover, the rational homotopy type of such a space is determined solely by its cohomology algebra (at least for a large class of such spaces).

3. $RCDGA$ AND G -FORMALITY

In this paper, we shall be concerned with equivariant versions of standard formality results. Let G be a compact, connected Lie group. Then $H^\bullet(BG; k)$ is isomorphic to the $kCDGA$ freely generated by a finite number of generators of even degree. We shall denote $R \stackrel{\text{def}}{=} H^\bullet(BG)$. We define the category $RCDGA^\circ$ to be the category of commutative differential graded R -algebras. We shall continue to assume that $H^0(A) \approx k$ for all algebras A . Thus, we obtain a faithful forgetful functor from $RCDGA^\circ$ to $kCDGA^\circ$. We also define $RCDGA$ to be the category of commutative differential graded R -algebras augmented over R . Composing augmentations with the augmentation $R \rightarrow k$, we get a faithful forgetful functor from $RCDGA$ to $kCDGA$.

It is a standard result that if \mathcal{C} is a closed model category and B is an object of \mathcal{C} , then the “over category” \mathcal{C}/B whose objects are maps $X \rightarrow B$ and whose morphisms are commutative squares of the type

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow & & \downarrow \\ B & \xlongequal{\quad} & B \end{array}$$

may be given the structure of a closed model category with the following definitions. Such a morphism in \mathcal{C}/B will be called a fibration, cofibration, or weak equivalence, if the map $f : X \rightarrow Y$ is such in \mathcal{C} . A similar statement holds for the “under category”, $B \backslash \mathcal{C}$. See [7] for these and other results about closed model categories.

Thus we see that both $RCDGA^\circ = R \backslash kCDGA^\circ$ and $RCDGA = RCDGA^\circ / R$ are closed model categories. Moreover, the simplicial category structure on $kCDGA^\circ$ defined in [4], section 5, induces a simplicial category structure on $RCDGA^\circ$ and $RCDGA$ in such a way that the results of [4], section 5, suitably modified, hold for these categories as well (cf. [22], II.2, proposition 6). From this, it follows that the homotopy results of [4], section 6, suitably modified, hold for $RCDGA^\circ$ and $RCDGA$ as well.

Definition 3.1. We shall say that an $RCDGA$ (resp. $RCDGA^\circ$) A is *formal* if $A \approx H(A)$ in $Ho(RCDGA)$ (resp. $Ho(RCDGA^\circ)$).

If a functor $j : \mathcal{C}_1 \rightarrow \mathcal{C}_2$ between two closed model categories preserves weak equivalences, then $X \approx Y$ in $Ho \mathcal{C}_1$ implies $j(X) \approx j(Y)$ in $Ho \mathcal{C}_2$. Thus if an algebra A is formal as an $RCDGA$, then it is formal as an $RCDGA^\circ$, and as a $kCDGA$, etc.

Suppose a smooth manifold M has a smooth action of a compact Lie group G . The equivariant cohomology of M may be computed by means of the Cartan complex of equivariant differential forms: $A_G^\bullet(M) = ((S\mathfrak{g}^* \otimes A^\bullet(M))^G, d_G)$ where the differential, d_G , is zero on $S\mathfrak{g}^*$, and for $\alpha \in A^\bullet(M)$, $d_G\alpha = d\alpha - \sum u_i \iota_{X_i} \alpha$, where

the $\{X_i\}$ are fundamental vector fields of the action corresponding to a basis of \mathfrak{g} , and the $\{u_i\}$ are the corresponding algebra generators of $S\mathfrak{g}^*$, which are given degree two. If M is just a topological space, we may compute the equivariant cohomology of M by means of the $\mathbb{Q}CDGA$ $A_G^\bullet(M)$ of [2], when $G = S^1$. Alternatively, we could use the de Rham algebra of the Borel construction, $A^\bullet(EG \times_G M)$ when M is a manifold, or the Sullivan-de Rham algebra of the Borel construction when M is not a manifold. We shall let $A_G^\bullet(M)$ possibly denote any of the above $kCDGA$'s, leaving it to the reader to interpret which model one prefers to use, as well as which ground field k . For a comprehensive treatment of equivariant de Rham theory, see [13].

Using either model, it is obvious how to obtain an R -algebra structure on $A_G^\bullet(M)$. It is induced by $R \xrightarrow{i} A_G^\bullet(pt.) \rightarrow A_G^\bullet(M)$, where the first map is a choice of minimal model for $A_G^\bullet(pt.)$ in $kCDGA$, and the second map is induced from the map $M \rightarrow \{pt.\}$. If we use the Cartan models, then the algebras $A_G^\bullet(M)$ are manifestly augmented over R when the group action has a fixed point. This is because in the Cartan model, $A_G^\bullet(pt.) = R$, and the inclusion of a fixed point gives a map $A_G^\bullet(M) \rightarrow A_G^\bullet(pt.) = R$. However, if we use the Sullivan-de Rham complex of the Borel construction, then $A_G^\bullet(pt.) = A^\bullet(BG) \neq R$. Thus we must use a quasi-isomorphic complex that is smaller and augmented over R . In [1], Allday shows that the complex $\eta^{-1}(R)$ is quasi-isomorphic to $A_G^\bullet(M)$, where $\eta : A_G^\bullet(M) \rightarrow A_G^\bullet(pt.)$ is induced by the inclusion of a fixed point into M , and R is embedded in $A_G^\bullet(pt.)$ via i as above. Clearly, $\eta^{-1}(R)$ is augmented over R , and is functorial for equivariant maps of G -spaces. We shall abuse notation and continue to write $A_G^\bullet(M)$, even when we may really mean $\eta^{-1}(R)$.

Let $GTOP$ denote the category of path-connected topological G -spaces with morphisms the equivariant maps. Then the under category $\{pt.\} \backslash GTOP$ consists of “based G -spaces”, which is the same thing as G -spaces with non-empty fixed-point set and a choice of base point in the fixed-point set. Then $A_G^\bullet(-)$ gives a functor from $GTOP$ to $RCDGA^\circ$ and from $\{pt.\} \backslash GTOP$ to $RCDGA$.

Definition 3.2. We shall say that a G -space M is *G -formal* if $A_G^\bullet(M)$ is formal as an $RCDGA^\circ$. A G -space M with equivariant base point p (i.e., a choice of fixed point $p \in M^G$) is *G -formal at p* if $A_G^\bullet(M)$ is formal as an $RCDGA$, where $A_G^\bullet(M)$ is augmented via the inclusion of p into M .

If we continue to define a minimal model of an $RCDGA$ as a connected cofibrant model \mathcal{M} for which the induced differential on $\ker \varepsilon / (\ker \varepsilon)^2$ is zero, where ε is an augmentation over R , then there may not be a minimal model for every algebra in $RCDGA$. An example is S^1 acting by rotations of S^2 about an axis. It is easy to see that there can be no minimal model for $A_{S^1}^\bullet(S^2)$ in $RCDGA$. However, there is a fairly canonical choice of cofibrant model for an $RCDGA$. Let $R \rightarrow A$ be an $RCDGA^\circ$. Then the map $R \rightarrow A$, viewed in $kCDGA$, may be factored as $R \rightarrow R \otimes_k \mathcal{M} \rightarrow A$ with the first map the inclusion, the latter map a quasi-isomorphism, and \mathcal{M} a minimal KS-complex, [14]. Note that the differential on $R \otimes \mathcal{M}$ is not the tensor product differential; see the definition of a KS-complex (Definition 2.1). The map $R \rightarrow R \otimes \mathcal{M}$ is a minimal KS-extension, in particular a cofibration in $kCDGA$, and hence we see that $R \otimes \mathcal{M}$ is a cofibrant model for A in $RCDGA^\circ$. Suppose A is, moreover, an algebra in $RCDGA$, and let $\varepsilon : A \rightarrow R$ be its augmentation. Then composing $R \otimes \mathcal{M} \rightarrow A \xrightarrow{\varepsilon} R$ gives an R -augmentation for

$R \otimes \mathcal{M}$. Thus, $R \otimes \mathcal{M}$ becomes a cofibrant model for A in the category $RCDGA$. As defined, it is unique up to isomorphism.

For those algebras of the form $A_G^\bullet(M)$ arising from a group action on the space M , this cofibrant model is more explicitly given by the Grivel-Halperin-Thomas theorem, which states that there is a commutative diagram

$$(2) \quad \begin{array}{ccccc} A_G^\bullet(pt.) & \longrightarrow & A_G^\bullet(M) & \longrightarrow & A^\bullet(M) \\ \uparrow & & \uparrow & & \uparrow \\ R & \xrightarrow{i} & R \otimes_k \mathcal{M} & \longrightarrow & \mathcal{M} \end{array}$$

associated to the fibration $M \rightarrow EG \times_G M \rightarrow BG$, where \mathcal{M} is a minimal model for M , and the bottom row is a Λ -minimal Λ -extension, see [12], [14].

Definition 3.3. We shall refer to $R \otimes \mathcal{M}$ as the G -model of A , or just simply as the G -model of M , when $A = A_G^\bullet(M)$.

Sometimes we may choose to denote it by $\mathcal{M}_G \stackrel{\text{def}}{=} R \otimes \mathcal{M}$. Note that $R \otimes \mathcal{M}$ may fail to be minimal as a $kCDGA$.

Following [1], [3], given a path-connected G -space M with equivariant base point (i.e., a fixed point) p , the equivariant (pseudo-dual) k -homotopy groups are defined to be

$$(3) \quad \Pi_{G,p}^n(M) \stackrel{\text{def}}{=} \pi^n(R \otimes \mathcal{M}) = H^n(\ker \varepsilon / (\ker \varepsilon)^2),$$

where $\varepsilon : R \otimes \mathcal{M} \rightarrow R$ is the R -algebra augmentation induced by the inclusion of p into M , as above. The assignment $(M, p) \mapsto (R \otimes \mathcal{M}, \varepsilon)$ gives a functor from $\{pt.\} \backslash \mathcal{GTOP}$ to $Ho(RCDGA)$, and the equivariant pseudo-dual k -homotopy groups are functorial as well. Note that if M is G -formal, then the equivariant pseudo-dual k -homotopy groups are determined by the equivariant cohomology ring of M .

The following lemma is useful for comparing the equivariant pseudo-dual k -homotopy groups to the ordinary pseudo-dual k -homotopy groups of the Borel construction.

Lemma 3.4. Let A be an $RCDGA$ and let $R \otimes \mathcal{M}$ be the G -model for A . Then $R \otimes \mathcal{M}$ is minimal in $kCDGA$.

Proof. We have the augmentation $\varepsilon : R \otimes \mathcal{M} \rightarrow R$, which is a map of $RCDGA$'s. The differential, D , on $R \otimes \mathcal{M}$ satisfies $D(r \otimes 1) = 0$, for $r \in R$, and generally has the form $D(1 \otimes \alpha) = r \otimes 1 + \sum r_i \otimes \alpha_i + 1 \otimes d\alpha$, where $\alpha, \alpha_i \in \mathcal{M}$, $r, r_i \in R$ with $\deg(\alpha), \deg(\alpha_i), \deg(r), \deg(r_i) > 0$, and where d is the differential in \mathcal{M} . Now, $0 = \varepsilon D(1 \otimes \alpha) = r + \sum r_i \varepsilon(\alpha_i) + \varepsilon(d\alpha)$. Since $d\alpha \in (\mathcal{M}^+)^2$, and ε is an algebra map, it follows that $\sum r_i \varepsilon(\alpha_i) + \varepsilon(d\alpha) \in (R^+)^2$. Hence, we must have that $r = 0$, and it follows that $R \otimes \mathcal{M}$ is minimal. \square

As an example, the pseudo-dual k -homotopy groups of the Borel construction of S^1 acting on S^2 do not distinguish the trivial action from a standard non-trivial one, whereas the equivariant pseudo-dual k -homotopy groups do.

4. GENERALITIES CONCERNING G -FORMALITY

In this section we give some basic results about G -formality, including reduction to subgroups and the G -formality of products and wedges.

We begin by noting that formality in the category $RCDGA^o$ is equivalent to formality in $kCDGA$. In general, for two R -algebras A and B , $A \approx B$ in $Ho(kCDGA)$ does not imply that $A \approx B$ in $Ho(RCDGA^o)$. Nevertheless, we have the following.

Lemma 4.1. *Assume that $R \xrightarrow{j} A$ is an $RCDGA^o$ and that we give $H(A)$ the R -algebra structure $R \xrightarrow{j^*} H(A)$. Then A is formal in $kCDGA$ if and only if A is formal in $RCDGA^o$.*

Proof. If A is formal in $RCDGA^o$, then it will be so in $kCDGA$, as we have noted above. Let us now assume that A is formal in $kCDGA$. Let \mathcal{N} be a minimal model for A and let $R \otimes \mathcal{M}$ be the G -model for A . Then we have a commutative diagram of $kCDGA$'s

$$(4) \quad \begin{array}{ccccc} R & \xrightarrow{j} & A & \xlongequal{\quad} & A \\ \parallel & & \uparrow \eta & & \uparrow \\ R & \longrightarrow & R \otimes \mathcal{M} & & \mathcal{N} \\ & & & & \downarrow \\ & & & & H(A) \end{array}$$

Since $R \otimes \mathcal{M}$ is cofibrant in $kCDGA$, there exists a map, which is necessarily a quasi-isomorphism, $R \otimes \mathcal{M} \rightarrow \mathcal{N}$ making the upper right square homotopy commute. This gives us a quasi-isomorphism $\varphi : R \otimes \mathcal{M} \rightarrow \mathcal{N} \rightarrow H(A)$. Then the map

$$(5) \quad R \otimes \mathcal{M} \xrightarrow{\varphi} H(A) \xrightarrow{(\varphi^*)^{-1}} H(R \otimes \mathcal{M}) \xrightarrow{\eta^*} H(A)$$

is a quasi-isomorphism and a map of R -algebras. \square

Remark 4.2. We note that this is not true for maps, however. That is, if $f : A \rightarrow B$ is a map of $RCDGA^o$'s, and f is formal as a map of $kCDGA$'s, then f need not be a formal map of $RCDGA^o$'s.

In the category $RCDGA$, formality is a concept distinct from formality in the category $kCDGA$. In fact, it is easy to see that M is G -formal at p if and only if the map $i : BG \rightarrow EG \times_G M$ is a formal map, where i is the map induced by the inclusion of p into M .

Definition 4.3. Suppose that G acts on a space M . Then the Serre spectral sequence associated with the fibration $M \rightarrow EG \times_G M \rightarrow BG$ is the same as the spectral sequence (from E_2 onwards) obtained from the G -model $R \otimes \mathcal{M}$ via the filtration $\mathcal{F}^p = R^{\geq p} \otimes \mathcal{M}$. If this spectral sequence degenerates at the E_2 term, then [11] refers to M as being equivariantly formal. For obvious reasons, we wish to avoid this terminology; however, to conform as well to current trends, we shall say that M is *ef* when this spectral sequence degenerates at the E_2 term.

Proposition 4.4. *Let G act on a space M . Suppose that $K \subset G$ is a closed, connected subgroup. If M is G -formal at p (or G -formal) and *ef*, then M is K -formal at p (resp. K -formal).*

Proof. We first consider the case where M is G -formal at p . The inclusion $K \subset G$ induces a pull-back diagram

$$(6) \quad \begin{array}{ccc} EK \times_K M & \longrightarrow & EG \times_G M \\ \downarrow & & \downarrow p \\ BK & \xrightarrow{i} & BG \end{array}$$

We shall denote $H^\bullet(BG)$ by R_G , and similarly for R_K . If we are using the Cartan complex of equivariant differential forms, then there is no problem with the proof. If we are using Allday's construction, $\eta^{-1}(R)$, as notated above, then we face the possibility that this construction may not be functorial with respect to changing the group. This is because there may not exist choices of minimal models so that $R_G \rightarrow R_K \rightarrow A^\bullet(BK)$ commutes with $R_G \rightarrow A^\bullet(BG) \xrightarrow{i} A^\bullet(BK)$. Then there would not exist an induced map $\eta^{-1}(R_G) \rightarrow \eta^{-1}(R_K)$.

This problem may be circumvented by the following procedure, as pointed out to us by C. Allday. Consider the diagram (6). Let \tilde{Y} denote the mapping cylinder of the top row, and Y the mapping cylinder of the bottom row. Then we have a commutative diagram

$$(7) \quad \begin{array}{ccccc} EK \times_K M & \xrightarrow{\tilde{j}_1} & \tilde{Y} & \xrightarrow{\tilde{j}_2} & EG \times_G M \\ \downarrow & & \downarrow & & \downarrow \\ BK & \xrightarrow{j_1} & Y & \xrightarrow{j_2} & BG \end{array}$$

in which the maps j_1, \tilde{j}_1 induce surjections on differential forms and the maps j_2, \tilde{j}_2 induce quasi-isomorphisms on differential forms. It is easy to show that we may use \tilde{Y} to form the complex $\eta^{-1}(R_G)$, as discussed in section 3, and that this complex will be quasi-isomorphic to $A_G^\bullet(M)$, and G -formal at p if M is G -formal at p . Moreover, we now may obtain a commutative diagram

$$(8) \quad \begin{array}{ccc} A^\bullet(Y) & \xrightarrow{j_1} & A^\bullet(BK) \\ \uparrow & & \uparrow \\ R_G & \xrightarrow{\tilde{j}_1} & R_K \end{array}$$

in which the vertical arrows are quasi-isomorphisms, since the map $A^\bullet(Y) \rightarrow A^\bullet(BK)$ is onto. This follows by the result for $RCDGA$, which is the analog of 6.9 in [4].

By Lemma A.1 of the appendix, there is a quasi-isomorphism of $kCDGA$'s

$$(9) \quad \bar{B}(A^\bullet(BK), A^\bullet(BG), A_G^\bullet(M)) \rightarrow A_K^\bullet(M),$$

where we are abusing notation in the event that we are using Allday's construction. Then, in either case, we obtain a quasi-isomorphism of $R_K CDGA$'s

$$(10) \quad \bar{B}(R_K, R_G, A_G^\bullet(M)) \rightarrow A_K^\bullet(M).$$

The bar complex (10) is an R_K -algebra via the R_K factor, and has an R_K -augmentation given by $\varepsilon(r_K, \alpha) = r_K \tilde{i}(\varepsilon_G(\alpha))$, where $r_K \in R_K$, $\alpha \in A_G^\bullet(M)$, and $\varepsilon_G :$

$A_G^\bullet(M) \rightarrow R_G$ is the augmentation of M for the action of G . By the assumption of G -formality, we get a commuting diagram whose vertical arrows are quasi-isomorphisms:

(11)

$$\begin{array}{ccccc} R_K & \longleftarrow & R_G & \longrightarrow & A_G^\bullet(M) \\ \parallel & & \parallel & & \uparrow \\ R_K & \longleftarrow & R_G & \longrightarrow & \mathcal{M}_G(M) \\ \parallel & & \parallel & & \downarrow \\ R_K & \longleftarrow & R_G & \longrightarrow & H_G^\bullet(M) \end{array}$$

Then we obtain the following sequence of maps, which are seen to be R_K CDGA quasi-isomorphisms by standard comparison theorems for their associated Eilenberg-Moore spectral sequences:

(12)

$$\bar{B}(R_K, R_G, A_G^\bullet(M)) \leftarrow \bar{B}(R_K, R_G, \mathcal{M}_G(M)) \rightarrow \bar{B}(R_K, R_G, H_G^\bullet(M)).$$

Now the bar complex $\bar{B}(R_K, R_G, H_G^\bullet(M))$ has only the single differential δ , and computes $\mathrm{Tor}_{R_G}(R_K, H_G^\bullet(M))$. Since M is ef, $H_G^\bullet(M)$ is a free R_G -module. Hence we have that $\bar{B}(R_K, R_G, H_G^\bullet(M))_\bullet$ is acyclic in bar degrees greater than zero, and the projection to cohomology

(13)

$$\begin{aligned} \bar{B}(R_K, R_G, H_G^\bullet(M))_\bullet &\rightarrow \bar{B}(R_K, R_G, H_G^\bullet(M))_0 \\ &\rightarrow H_\bullet(\bar{B}(R_K, R_G, H_G^\bullet(M))_\bullet) \\ &\approx H_K^\bullet(M) \end{aligned}$$

is an R_K CDGA quasi-isomorphism.

The case where we consider M to be G -formal in the category $R_G\mathrm{CDGA}^o$ is similar. □

Corollary 4.5. *Let G act on a space M . Suppose that M is G -formal at p (or G -formal) and ef. Then M is formal in $k\mathrm{CDGA}$.*

Proof. Just take K to be the identity subgroup in Proposition 4.4. □

Remark 4.6. If we use Remark A.5 of the appendix, then we can see that a K -model for M is given by $\bar{B}_{R_G}(R_K, R_G, \mathcal{M}_G(M)) = R_K \otimes_{R_G} \mathcal{M}_G(M)$.

In line with the general theme of considering maximal tori in compact, connected Lie groups, we have the following fact, which is due to C. Allday. We say that a space M is of *finite type* if $H^i(M)$ is a finite-dimensional k -vector space for all i .

Proposition 4.7. *Let G act on M , and let $T \subset G$ be a maximal torus. If M is G -formal (G -formal at p), then M is T -formal (resp. T -formal at p). Moreover, if M is a space of finite type, $p \in M^G$, and M is T -formal at p , then M is G -formal at p .*

Proof. We can already see that G -formal implies T -formal by the proof of Proposition 4.4. We only need the fact that now R_T is a free R_G -module, which follows from the well-known fact that as R_G -modules, $R_T \approx R_G \otimes H^\bullet(G/T)$.

Showing that T -formal at p implies G -formal at p may be achieved by imitating the proof that $A \otimes K$ being formal implies A is formal, for K an extension field of k , which is corollary 6.9 of [15]. We omit the details, but mention the setup. First,

we see by Remark 4.6 that a T -model for $A_T^\bullet(M)$ is given by $R_T \otimes_{R_G} \mathcal{M}_G(M)$ with differential $1 \otimes D_G$, where D_G is the differential for the G -model $\mathcal{M}_G(M)$. Thus it suffices to show that if $R_T \otimes_{R_G} \mathcal{M}_G(M)$ is formal as an R_T CDGA, then $\mathcal{M}_G(M)$ is formal as an R_G CDGA.

It turns out that the constructions of bigraded and filtered models of the relevant algebras, as in [15], give models in the category $RCDGA$. The proof of corollary 6.9 may be imitated without too much difficulty. \square

Proposition 4.8. *Suppose that X and Y are G -spaces, both of which are G -formal (or assume X is G -formal at p and Y is G -formal at q), and suppose that one or both of them is ef. Then $X \times Y$ is G -formal (resp. G -formal at (p, q)) for the diagonal action of G .*

Proof. The pull-back diagram

$$(14) \quad \begin{array}{ccc} X \times Y & \longrightarrow & Y \\ \downarrow & & \downarrow \\ X & \longrightarrow & \{pt.\} \end{array}$$

gives rise to a pull-back diagram

$$(15) \quad \begin{array}{ccc} EG \times_G (X \times Y) & \longrightarrow & EG \times_G Y \\ \downarrow & & \downarrow \\ EG \times_G X & \longrightarrow & BG \end{array}$$

Then we obtain an $RCDGA^\circ$ quasi-isomorphism

$$(16) \quad \bar{B}(A_G^\bullet(X), A_G^\bullet(\{pt.\}), A_G^\bullet(Y)) \xrightarrow{\theta} A_G^\bullet(X \times Y)$$

by Lemma A.3 of the appendix. If X and Y both have fixed points, then so will their product $X \times Y$. In that case, θ is a quasi-isomorphism of $RCDGA$'s by Lemma A.3 of the appendix. Furthermore,

$$(17) \quad \bar{B}(A_G^\bullet(X), R, A_G^\bullet(Y)) \rightarrow \bar{B}(A_G^\bullet(X), A_G^\bullet(\{pt.\}), A_G^\bullet(Y))$$

is an $RCDGA^\circ$ ($RCDGA$) quasi-isomorphism. Since X and Y are G -formal, we get $RCDGA^\circ$ ($RCDGA$) quasi-isomorphisms of bar complexes

$$(18) \quad \begin{array}{c} \bar{B}(A_G^\bullet(X), R, A_G^\bullet(Y)) \\ \uparrow \\ \bar{B}(R \otimes \mathcal{M}(X), R, R \otimes \mathcal{M}(Y)) \\ \downarrow \\ \bar{B}(H_G^\bullet(X), R, H_G^\bullet(Y)) \end{array}$$

by standard arguments comparing the associated Eilenberg-Moore spectral sequences.

Since one or both of X, Y is ef, just as in the proof of Proposition 4.4, the bar complex $(\bar{B}(H_G^\bullet(X), R, H_G^\bullet(Y))_\bullet; \delta)$ is acyclic in degrees greater than zero with respect to the bar grading, and the projection to its cohomology is an $RCDGA^\circ$ ($RCDGA$) quasi-isomorphism. \square

Proposition 4.9. *Let X and Y be G -spaces whose fixed-point sets are non-empty. Picking base points $p \in X^G$ and $q \in Y^G$, we may form the wedge $X \vee Y$ along these base points. Then G acts on $X \vee Y$. If X is G -formal at p and Y is G -formal at q , then $X \vee Y$ is G -formal at the join of p and q .*

Proof. Let $\varepsilon_X, \varepsilon_Y$ denote the augmentations of equivariant differential forms, and let i_X, i_Y denote the inclusions of X, Y into $X \vee Y$. Then Mayer-Vietoris gives a short exact sequence

$$(19) \quad 0 \rightarrow A_G^\bullet(X \vee Y) \xrightarrow{i_X + i_Y} A_G^\bullet(X) \oplus A_G^\bullet(Y) \xrightarrow{\varepsilon_X - \varepsilon_Y} R \rightarrow 0.$$

Thus $i_X + i_Y$ induces an isomorphism

$$A_G^\bullet(X \vee Y) \approx A_G^\bullet(X) \bigoplus_R A_G^\bullet(Y) = \ker\{\varepsilon_X - \varepsilon_Y\}.$$

Moreover, since ε_X , say, induces a surjection in cohomology, the associated long exact sequence splits into short exact sequences, and thus

$$(20) \quad H_G^\bullet(X \vee Y) \approx H_G^\bullet(X) \oplus_R H_G^\bullet(Y).$$

Since X and Y are G -formal, we have maps $A_G^\bullet(X) \leftarrow \mathcal{M}_G(X) \rightarrow H_G^\bullet(X)$ which are quasi-isomorphisms of R CDGA's, and similarly for Y . So we have a commutative diagram whose rows are short exact sequences:

$$(21) \quad \begin{array}{ccccccccc} 0 & \longrightarrow & A_G^\bullet(X) \oplus_R A_G^\bullet(Y) & \longrightarrow & A_G^\bullet(X) \oplus A_G^\bullet(Y) & \longrightarrow & R & \longrightarrow & 0 \\ & & \uparrow & & \uparrow & & \parallel & & \parallel \\ 0 & \longrightarrow & \mathcal{M}_G(X) \oplus_R \mathcal{M}_G(Y) & \longrightarrow & \mathcal{M}_G(X) \oplus \mathcal{M}_G(Y) & \longrightarrow & R & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \parallel & & \parallel \\ 0 & \longrightarrow & H_G^\bullet(X) \oplus_R H_G^\bullet(Y) & \longrightarrow & H_G^\bullet(X) \oplus H_G^\bullet(Y) & \longrightarrow & R & \longrightarrow & 0 \end{array}$$

Then we obtain maps between the associated long exact sequences in cohomology. By the 5-lemma, it follows that the maps

$$(22) \quad A_G^\bullet(X) \oplus_R A_G^\bullet(Y) \leftarrow \mathcal{M}_G(X) \oplus_R \mathcal{M}_G(Y) \rightarrow H_G^\bullet(X) \oplus_R H_G^\bullet(Y)$$

are quasi-isomorphisms. It is easy to check that these maps are compatible with augmentations and the R -algebra structure, so are R CDGA quasi-isomorphisms. \square

5. EXAMPLES OF G -FORMAL SPACES

In this section we give some examples of G -formal spaces.

5.1. Compact Kähler manifolds. Let M be a compact Kähler manifold, and G a compact, connected Lie group acting on M by holomorphic transformations. We introduce equivariant holomorphic cohomology groups. Since M is a complex manifold, the complex-valued differential forms on M are bigraded in the usual way. We shall denote $S\mathfrak{g}^* \otimes_{\mathbb{R}} \mathbb{C}$ by simply $S\mathfrak{g}^*$. Then we define the equivariant Dolbeault cohomology to be the cohomology of the complex

$$(23) \quad ([S\mathfrak{g}^* \otimes_{\mathbb{C}} A^{p,\bullet}(M)]^G; \bar{\partial} + \sum u_i \iota_{Z_i}).$$

Here Z_i is the holomorphic vector field on M which comes about by splitting the fundamental vector field $X_i = Z_i + \bar{Z}_i$ into its holomorphic and anti-holomorphic

components. The generators $u_i \in S\mathfrak{g}^*$ are given bidegree $(1, 1)$. The operators act in a similar way as for the ordinary equivariant cohomology. We shall denote the q th cohomology of this complex by $H_G^{p,q}(M)$.

The following theorem was proved in [17] and independently established in [25].

Theorem 5.1. *Suppose that M is a compact Kähler manifold endowed with a holomorphic action of a compact, connected Lie group G , and suppose that M is ef for the action of G . Then M is G -formal. If $M^G \neq \emptyset$, then M is G -formal at any fixed point.*

Proof. The Cartan complex is $(A_G^\bullet(M), d_G) = ([S\mathfrak{g}^* \otimes A^\bullet(M; \mathbb{C})]^G; d + \sum u_i \iota_{X_i})$. Let $X_i = Z_i + \bar{Z}_i$ be the splitting of the fundamental vector field X_i into its holomorphic and anti-holomorphic parts. The differential $d = \partial + \bar{\partial}$ also splits. Hence we may split the equivariant differential as $d + \sum u_i \iota_{X_i} = (\bar{\partial} + \sum u_i \iota_{Z_i}) + (\partial + \sum u_i \iota_{\bar{Z}_i})$. The complex $[S\mathfrak{g}^* \otimes A_G^\bullet(M; \mathbb{C})]^G$ is bigraded by giving $u_i \in S\mathfrak{g}^*$ bidegree $(1, 1)$, and taking the usual bigrading on $A^\bullet(M; \mathbb{C})$. It is easy to show that

$$(24) \quad ([S\mathfrak{g}^* \otimes A^{\bullet,\bullet}(M; \mathbb{C})]^G; (\bar{\partial} + \sum u_i \iota_{Z_i}), (\partial + \sum u_i \iota_{\bar{Z}_i}))$$

is a first quadrant double complex. Accordingly we have two canonical filtrations of this complex. We claim that the spectral sequences corresponding to both of them degenerate at the E_1 term, and moreover are n -opposite, meaning that $'F^p \oplus ''F^q \approx H^n$ for $p + q - 1 = n$. Formality for $A_G^\bullet(M)$ then follows owing to the results in [6], sections 5 and 6.

Let us consider the filtration in which we take $\bar{\partial} + \sum u_i \iota_{Z_i}$ cohomology first. This is the Dolbeault equivariant cohomology defined above. It itself forms a first quadrant double complex with the two differentials $\bar{\partial}$ and $\sum u_i \iota_{Z_i}$. Let us filter so that we take the $\bar{\partial}$ cohomology first. Then the E_1 term for the equivariant Dolbeault complex is (additively)

$$(25) \quad H([S\mathfrak{g}^* \otimes A^\bullet(M)]^G; \bar{\partial}) \approx [H(S\mathfrak{g}^* \otimes A^\bullet(M); \bar{\partial})]^G \approx (S\mathfrak{g}^* \otimes H_{\bar{\partial}}^\bullet(M))^G$$

$$(26) \quad \approx (S\mathfrak{g}^*)^G \otimes H_{\bar{\partial}}^\bullet(M) \approx H^\bullet(BG) \otimes H_{\bar{\partial}}^\bullet(M).$$

Now by ordinary Hodge theory for compact Kähler manifolds, this last is isomorphic to $H^\bullet(BG) \otimes H^\bullet(M)$. But now there can be no further non-trivial differentials in the spectral sequence, by the assumption that M is ef. This result follows analogously for the other filtration, which is just the complex conjugate of this one. Furthermore, it is easy to see that the two filtrations are n -opposite.

Hence we have a “ $\partial_G \bar{\partial}_G$ -lemma” for the equivariant differential forms, where we mean by $\bar{\partial}_G$ the equivariant Dolbeault operator as defined above. Formality follows via the sequence of CCDGA quasi-isomorphisms

$$(27) \quad A_G^\bullet(M) \hookrightarrow \ker(\bar{\partial}_G) \rightarrow H_G^\bullet(M),$$

which are the inclusion and projection, respectively. These maps are maps of R -algebras, and moreover, it follows that for equivariant holomorphic maps between M and N , we get a commutative diagram linking the sequence (27) for M to the analogous sequence for N . In particular, if the action of G on M has fixed points, then the inclusion of one (chosen as an equivariant base point) gives augmentations so that the sequence (27) commutes with augmentations. That is, M is G -formal in $RCDGA$. \square

Corollary 5.2. *Suppose that M is a compact Kähler manifold endowed with a holomorphic action of a compact, connected Lie group G . Assume that $M^G \neq \emptyset$. Then M is G -formal at any fixed point.*

Proof. Let $p \in M^G$. Let $T \subset G$ be a maximal torus. Then $M^T \neq \emptyset$, and a theorem of Blanchard says that M is ef for the action of T ; see [9], Chapter XII, theorem 6.2. By Theorem 5.1, M is T -formal at p . By Proposition 4.7, M is G -formal at p . \square

Remark 5.3. The proof of Theorem 5.1 implies an equivariant Hodge decomposition

$$(28) \quad H_G^n(M) \approx \bigoplus_{p+q=n} H_G^{p,q}(M).$$

5.2. Elliptic spaces. We recall that an *elliptic space* M is a space such that both $H^\bullet(M; k)$ and V are finite-dimensional k -vector spaces, where $\mathcal{M}(M) = \bigwedge(V)$ is a minimal model for M . We shall use the following result of [19].

Proposition 5.4 (Lupton). *Let $F \rightarrow E \rightarrow B$ be a fibration in which F is formal and elliptic, and B is formal and simply-connected. If the Serre spectral sequence of the fibration degenerates at the E_2 term, then E is formal also.*

Theorem 5.5. *Let M be an elliptic G -space. If M is formal and ef, then M is G -formal. If $M^G \neq \emptyset$, then M is G -formal at any fixed point.*

Proof. We have the fibration $M \rightarrow EG \times_G M \rightarrow BG$. Then Proposition 5.4 implies that $A_G^\bullet(M)$ is formal as a k CDGA. The proof of Lupton's proposition works (adapting to our situation) by finding a model for $A_G^\bullet(M)$ of the form $R \otimes \mathcal{M}$ that is bigraded as a k CDGA. Here, \mathcal{M} is the bigraded (minimal) model of M . Elements of R are in degree zero for the second grading, so that $(R \otimes \mathcal{M})_0 = R \otimes (\mathcal{M})_0$. It is shown that with respect to the second grading we have $H_+(R \otimes \mathcal{M}) = 0$, and hence the projection to cohomology

$$(29) \quad R \otimes \mathcal{M} \rightarrow (R \otimes \mathcal{M})_0 \rightarrow H_G^\bullet(M)$$

is a quasi-isomorphism. Clearly, this is a map of R -algebras. Moreover, if $M^G \neq \emptyset$, then this map commutes with the augmentations over R . This follows because first the map $R \otimes \mathcal{M} \rightarrow (R \otimes \mathcal{M})_0$ commutes with augmentations. Second, since the augmentation $\varepsilon : R \otimes \mathcal{M} \rightarrow R$ is a map of R CDGA's, $\varepsilon(d\alpha) = 0$ for all α , so that the map $(R \otimes \mathcal{M})_0 \rightarrow H_G^\bullet(M)$ commutes with augmentations. \square

Corollary 5.6. *Let M be an elliptic space. Suppose that a torus T acts on M with $M^T \neq \emptyset$. Suppose further that one of the components of the fixed-point set, say M_i^T , satisfies $H^{\text{odd}}(M_i^T) = 0$. Then M is T -formal at any fixed point.*

Proof. Since M is elliptic, it follows (via localization and localization for equivariant rational homotopy [3]) that each component of the fixed-point set is elliptic and $\chi_\pi(M) = \chi_\pi(M_i^T)$, where χ_π is the homotopy Euler characteristic. But Halperin has shown that for elliptic spaces the conditions $H^{\text{odd}} = 0$ and $\chi_\pi = 0$ are equivalent, and moreover such spaces are formal. Thus $0 = \chi_\pi(M_i^T) = \chi_\pi(M)$. Hence M is formal and $H^{\text{odd}}(M) = 0$. But this latter condition implies that M is ef. So we may apply Theorem 5.5. \square

Remark 5.7. Suppose G acts on a simply-connected space M with non-empty fixed-point set. Then by picking a base point in the fixed-point set, we obtain an action

of G on the space of based loops in M , denoted ΩM . Since the cohomology of ΩM is free, we see that ΩM will be G -formal if ΩM is ef. (Lupton's proof could be extended to this case, as well.) If $G = T$ is a torus, and M is elliptic, then the condition that ΩM is ef is equivalent to the G -model $R \otimes \mathcal{M}(M)$ being minimal in the category $RCDGA$; see [3], 3.3.15.

5.3. Miscellanea. Next we shall give a few extra examples of G -formality.

Theorem 5.8. *Let M be a space with minimal model $\mathcal{M} = \bigwedge(V)$. Suppose that $dx = 0$ for all $x \in V^{even}$ such that $\deg(x) < \dim M$. Suppose further that the circle $S^1 = T$ acts on M , that M is ef, and that each component of the fixed-point set is formal and satisfies $H^{odd}(M^T) = 0$. Then M is T -formal at any fixed point.*

Proof. Since M is ef, the Serre spectral sequence for the fibration $M \rightarrow ET \times_T M \rightarrow BT$ degenerates at the E_2 term. (Note that by the localization theorem, this implies that $M^T \neq \emptyset$.) By a standard change of basis argument, we may assume that in the T -model $(R \otimes \mathcal{M}, D)$ we have $Dx = 0$, for $x \in V^{even}$ such that $\deg(x) < \dim M$. Let $i : M^T \hookrightarrow M$ denote the inclusion of the fixed-point set. Then we have maps of $RCDGA$'s (actually, the algebras on the right-hand side of the diagram do not satisfy $H^0 = k$, but this will not present any problems)

$$\begin{array}{ccc}
 A_T^\bullet(M) & \xrightarrow{i} & A_T^\bullet(M^T) \\
 \uparrow & & \uparrow \\
 R \otimes \mathcal{M}(M) & \xrightarrow{i} & R \otimes \mathcal{M}(M^T) \\
 & & \downarrow h \\
 H_T^\bullet(M) & \xrightarrow{i^*} & H_T^\bullet(M^T) = R \otimes H^\bullet(M^T)
 \end{array}
 \tag{30}$$

where h is a quasi-isomorphism since M^T is formal. Since M is ef, the map i^* is an injection. We claim that $hi(R \otimes \mathcal{M}(M)) \subseteq i^*(H_T^\bullet(M))$. Since the maps are algebra maps, it suffices to check this on algebra generators. Since M is ef, the localization theorem shows that i^* is an isomorphism in degrees $\geq \dim M$. Also if $\alpha \in R \otimes \mathcal{M}(M)$ has odd degree, then $hi(\alpha) = 0$, since $H_T^{odd}(M^T) = 0$ by assumption. So it suffices to check the claim on algebra generators of $R \otimes \mathcal{M}(M)$ of even degree less than $\dim M$. Let α be such a generator. If $\alpha \in R$, then the claim is obviously true. If $\alpha \in \mathcal{M}(M)$, then by assumption $D\alpha = 0$. Then $hi(\alpha) = [i(\alpha)] = j([\alpha])$. Thus we have a map

$$R \otimes \mathcal{M}(M) \xrightarrow{hi} H_T^\bullet(M),
 \tag{31}$$

which is a quasi-isomorphism of $RCDGA$'s. \square

Corollary 5.9. *Let M be a simply-connected space with minimal model $\bigwedge(V)$. Suppose that $dx = 0$ for all $x \in V^{even}$ such that $\deg(x) < \dim M$. Suppose further that a torus T acts on M , that M is ef, and that each component of the fixed-point set is formal and satisfies $H^{odd}(M^T) = 0$. Then M is formal in $kCDGA$.*

Proof. First of all, there is a subcircle $S^1 \subset T$ such that $M^{S^1} = M^T$. The inclusion of this circle $S^1 \hookrightarrow T$ induces a pull-back diagram:

$$(32) \quad \begin{array}{ccc} ES^1 \times_{S^1} M & \longrightarrow & ET \times_T M \\ \downarrow & & \downarrow \\ BS^1 & \longrightarrow & BT \end{array}$$

Since the action of T is ef, the Serre spectral sequence for the fibration on the right degenerates at the E_2 term. But then the same is true for the pull-back fibration. Hence the S^1 action is ef as well. Now the result follows from Theorem 5.8 and Corollary 4.5. \square

Corollary 5.10. *Let M^4 be a space such that $H^{\text{odd}}(M) = 0$ and $\dim M = 4$. Suppose that a circle $S^1 = T$ acts on M . Then M is T -formal at any fixed point.*

Proof. We have that $H^{\text{odd}}(M) = 0$, so that M is ef. Then $M^T \neq \emptyset$. By localization, $H^{\text{odd}}(M^T) = 0$. But path-connected spaces with $H^1 = 0$ of dimension less than or equal to 4 are formal; so each component of M^T is formal. The result follows by Theorem 5.8. \square

Remark 5.11. A simple example of an S^1 -space satisfying the conditions of Theorem 5.8, but which is not Kähler or elliptic, is the following. Let S^1 act on S^4 so that the fixed-point set consists of two isolated points. Extend this to a diagonal action of S^1 on $S^4 \times S^4$. Then, removing a neighborhood of a fixed point, we may form the connected sum $S^4 \times S^4 \# S^4 \times S^4$. This manifold then inherits an S^1 action with 6 isolated fixed points. It is not elliptic, and not even symplectic, since $H^2 = 0$. It is easy to check that it satisfies the conditions of Theorem 5.8, so is S^1 -formal. (This can also be seen by proving that the connected sum (made in an equivariant setting) of G -formal spaces is again G -formal, which we have omitted.)

We conclude this section with two examples that do not involve the condition of M being ef.

Lemma 5.12. *Let M be a simply-connected compact manifold. Suppose that G acts freely on M and $\dim G \geq \dim M - 6$. Then M is G -formal.*

Proof. Since G acts freely, M/G is a simply-connected manifold of dimension 6 or less. Hence M/G is formal [21]. So $EG \times_G M$ is formal. \square

Remark 5.13. Suppose, in the situation of Lemma 5.12, we have that $\dim M - 6 > \text{rank}(G)$. Let $T \subset G$ be a maximal torus. Then by Proposition 4.7, M/T is a simply-connected manifold of dimension greater than 6 which is formal.

Lemma 5.14. *Let M be a simply-connected elliptic space. Suppose that G acts almost freely on M (meaning all isotropy groups are finite), and $\text{rank}(G) = -\chi_\pi(M)$. Then M is G -formal.*

Proof. Since M and BG have finite-dimensional pseudo-dual rational homotopy, so does $EG \times_G M$, as may be seen by considering the fibration $M \rightarrow EG \times_G M \rightarrow BG$. Since G acts almost freely, $H^\bullet(EG \times_G M)$ is finite-dimensional as well. Furthermore,

$$(33) \quad \chi_\pi(EG \times_G M) = \chi_\pi(M) + \chi_\pi(BG) = -\text{rank}(G) + \text{rank}(G) = 0.$$

Thus $EG \times_G M$ is elliptic with $\chi_\pi = 0$, so is formal. \square

6. AN APPLICATION

In this section we give an application of G -formality. We will show that the computation of the equivariant cohomology of loop spaces simplifies considerably when the space is G -formal.

Let us consider a simply-connected space M . Suppose that G acts on M with non-empty fixed-point set. Let $p \in M^G$ be a choice of base point. Then we get an action of G on the loops in M based at p , $\Omega(M; p)$, which we shall often abbreviate as ΩM . Let $P(M; p)$ be the space of paths in M , based at p . Then we have the fibration

$$(34) \quad \begin{array}{ccc} \Omega M & \longrightarrow & P(M; p) \\ \downarrow & & \downarrow \pi \\ \{p\} & \longrightarrow & M \end{array}$$

where π is the map sending a path $\gamma(t)$ to its value at time 1, $\gamma(1)$. Moreover, the G -action induces a pull-back diagram of fibrations

$$(35) \quad \begin{array}{ccc} EG \times_G \Omega M & \longrightarrow & EG \times_G P(M; p) \\ \downarrow & & \downarrow \pi \\ BG & \longrightarrow & EG \times_G M \end{array}$$

Hence there is a quasi-isomorphism of $RCDGA$'s

$$(36) \quad \theta : \bar{B}(R, A_G^\bullet(M), A_G^\bullet(P(M; p))) \rightarrow A_G^\bullet(\Omega(M))$$

by Lemma A.3 of the appendix. Now the inclusion of $\{p\}$ into $P(M; p)$ followed by π is the inclusion of $\{p\}$ into M . These maps are equivariant, so induce their analogs on the Borel constructions. Hence we get an $RCDGA$ quasi-isomorphism

$$(37) \quad \bar{B}(R, A_G^\bullet(M), R) \leftarrow \bar{B}(R, A_G^\bullet(M), A_G^\bullet(P(M; p))).$$

Proposition 6.1. *Let G act on a simply-connected space M with non-empty fixed-point set, so that G acts on ΩM . Suppose that M is G -formal. Then there is an isomorphism of R -algebras*

$$(38) \quad H_G^\bullet(\Omega M) \approx \text{Tor}_{H_G^\bullet(M)}(R, R).$$

Proof. We have that $A_G^\bullet(\Omega M)$ is quasi-isomorphic to $\bar{B}(R, A_G^\bullet(M), R)$ (via a sequence of $RCDGA$ quasi-isomorphisms). The assumption of G -formality means we have a commuting diagram of R -algebras

$$(39) \quad \begin{array}{ccccc} R & \xleftarrow{\varepsilon} & A_G^\bullet(M) & \xrightarrow{\varepsilon} & R \\ \parallel & & \uparrow & & \parallel \\ R & \xleftarrow{\varepsilon} & \mathcal{M}_G(M) & \xrightarrow{\varepsilon} & R \\ \parallel & & \downarrow & & \parallel \\ R & \xleftarrow{\varepsilon} & H_G^\bullet(M) & \xrightarrow{\varepsilon} & R \end{array}$$

We obtain $RCDGA$ quasi-isomorphisms

$$(40) \quad \bar{B}(R, A_G^\bullet(M), R) \leftarrow \bar{B}(R, \mathcal{M}_G(M), R) \rightarrow \bar{B}(R, H_G^\bullet(M), R).$$

This follows by standard comparison theorems for the Eilenberg-Moore spectral sequences associated to the bar complexes. Thus we have that $\bar{B}(R, H_G^\bullet(M), R)$ is quasi-isomorphic to $A_G^\bullet(\Omega M)$ (via a sequence of RCDGA quasi-isomorphisms). But the cohomology of $\bar{B}(R, H_G^\bullet(M), R)$ is $\text{Tor}_{H_G^\bullet(M)}(R, R)$. \square

Remark 6.2. We can always choose any resolution to compute Tor . But we note that we may always use the bar resolution, and using Lemma A.4 of the appendix, we see that when M is G -formal, $H_G^\bullet(\Omega M)$ may be computed via the (single) complex

$$(41) \quad (\bar{B}_R(R, H_G^\bullet(M), R); \delta).$$

Remark 6.3. We could also obtain analogous results for the equivariant cohomology of the free loop space LM .

7. AN EXAMPLE

In this section we compute an example of the equivariant cohomology of the based-loop space using the normalized bar complex over R of Remark 6.2.

7.1. Example: S^1 acting on ΩS^2 . The circle S^1 acts on the 2-sphere S^2 by rotations about an axis, say the z -axis when S^2 is the unit sphere in \mathbb{R}^3 . This action is holomorphic and Hamiltonian. Thus by Theorem 5.1, S^2 is G -formal ($G = S^1$). It is easy to show that the equivariant cohomology ring is

$$(42) \quad H_{S^1}^\bullet(S^2; k) \approx k[x, u]/(x + u)(x - u),$$

where the degree of x and u is two, and $R = k[u]$ acts as multiplication by u .

The fixed-point set, F , consists of the north and south poles. We shall write $F = \{N, S\}$. Let ΩS^2 be loops based at the north pole. Then S^1 acts on ΩS^2 . Then the equivariant cohomology of the based loops, $H_{S^1}(\Omega S^2)$, may be computed as the cohomology of the bar complex

$$(43) \quad (\bar{B}_{k[u]}(k[u], \frac{k[x, u]}{(x^2 - u^2)}, k[u]); \delta).$$

Let ω be the symplectic form on S^2 . Then x is represented by the form $\omega - uf \in A_{S^1}^\bullet(S^2)$, and u is represented by the form $u \in A_{S^1}^\bullet(S^2)$, using the Cartan complex of equivariant differential forms. Here, f is the moment map which sends a point on $S^2 \subset \mathbb{R}^3$ to its z -component. Then the inclusion of the north pole $\{N\}$ into S^2 induces the augmentation $H_{S^1}^\bullet(S^2) \rightarrow H_{S^1}^\bullet(\{N\}) \approx k[u]$ sending $x \mapsto -u$ and $u \mapsto u$. We omit the details of computing the bar complex, but one finds without difficulty the cohomology generators $(1, \underbrace{x, \dots, x}_n, 1)$ in degree n for n odd,

and $(u^{n/2}, 1)$ in degree n for n even. Owing to the shuffle product structure on the bar complex, one sees that, as an R -algebra,

$$(44) \quad H_{S^1}^\bullet(\Omega S^2) \approx k[u] \oplus \bigwedge(x_1) \oplus \bigwedge(x_3) \oplus \bigwedge(x_5) \dots,$$

where x_i is an indeterminate of degree i .

Remark 7.1. In this example, the normalized bar complex $\bar{B}_R(R, H_{S^1}^\bullet(S^2), R)$ is actually isomorphic to the k CDGA minimal model for $ES^1 \times_{S^1} \Omega S^2$, which is

$$(45) \quad \mathcal{M}_{ES^1 \times_{S^1} \Omega S^2} = \bigwedge(u, x, y) \quad (du = 0 = dx, dy = ux),$$

where the degrees of u and y are 2, and the degree of x is 1. The isomorphism is given by $(1, x, 1) \mapsto x$, $(u, 1) \mapsto u$, and $(1, x, x, 1) \mapsto y$.

Remark 7.2. In this example, the space $ES^1 \times_{S^1} \Omega S^2$ is not formal, implying that ΩS^2 is not G -formal. Indeed, Massey products abound.

APPENDIX A. BAR COMPLEXES AND EILENBERG-MOORE THEORY

In this appendix we shall discuss the theory of Eilenberg and Moore concerning pull-backs of fibrations. We will also consider equivariant versions of these results. For references, see [20], [24], or [8].

Let us suppose that we have a fibration $F \rightarrow E \xrightarrow{p} B$ and a map $f : X \rightarrow B$, so that we obtain a pull-back diagram:

$$(46) \quad \begin{array}{ccc} E_f & \xrightarrow{\tilde{f}} & E \\ \tilde{p} \downarrow & & \downarrow p \\ X & \xrightarrow{f} & B \end{array}$$

Then the maps f^* and p^* make $A^\bullet(X)$ and $A^\bullet(E)$ (differential graded) modules over $A^\bullet(B)$. Let us assume that B is simply-connected. Then a theorem of Eilenberg and Moore asserts that there is an isomorphism

$$(47) \quad \theta : \mathrm{Tor}_{A^\bullet(B)}(A^\bullet(X), A^\bullet(E)) \rightarrow H^\bullet(E_f).$$

We may use the bar resolution to obtain a resolution of, say, $A^\bullet(X)$ by $A^\bullet(B)$ -modules. Since we are considering $A^\bullet(-)$ to be the de Rham or Sullivan-de Rham complex, we will use Chen's normalized bar resolution, see [5] or [10].

More specifically, the bar complex is

$$(48) \quad B_k(A^\bullet(X), A^\bullet(B), A^\bullet(E)) = \bigoplus_{i=0}^{\infty} A^\bullet(X) \otimes_k (sA^\bullet(B))^{\otimes i} \otimes_k A^\bullet(E),$$

where the tensor products are over the ground field k , and s denotes the suspension functor on graded vector spaces that lowers the degree by one. Hence the degree of an element $(\alpha, \omega_1, \dots, \omega_k, \beta)$ is $\deg(\alpha) + \sum_{i=1}^k (\deg(\omega_i) - 1) + \deg(\beta)$, where $\alpha \in A^\bullet(X)$, $\omega_i \in A^\bullet(B)$, and $\beta \in A^\bullet(E)$. Actually, the bar complex is bigraded. We introduce the *bar degree*, denoted $B_k(A^\bullet(X), A^\bullet(B), A^\bullet(E))_\bullet$. The bar degree of an element $(\alpha, \omega_1, \dots, \omega_k, \beta)$ is defined to be $-k$. The other grading is the normal tensor product grading, the degree of an element $(\alpha, \omega_1, \dots, \omega_k, \beta)$ being $\deg(\alpha) + \sum_{i=1}^k \deg(\omega_i) + \deg(\beta)$.

There are two differentials of total degree $+1$:

$$(49) \quad d(\alpha, \omega_1, \dots, \omega_k, \beta) = (d\alpha, \omega_1, \dots, \omega_k, \beta) \\ + \sum_{i=1}^k (-1)^{\varepsilon_{i-1}+1} (\alpha, \omega_1, \dots, \omega_{i-1}, d\omega_i, \omega_{i+1}, \dots, \omega_k, \beta) \\ + (-1)^{\varepsilon_k} (\alpha, \omega_1, \dots, \omega_k, d\beta),$$

$$(50) \quad -\delta(\alpha, \omega_1, \dots, \omega_k, n) = (-1)^{\varepsilon_0} (\alpha\omega_1, \omega_2, \dots, \omega_k, \beta) \\ + \sum_{i=1}^{k-1} (-1)^{\varepsilon_i} (\alpha, \omega_1, \dots, \omega_{i-1}, \omega_i\omega_{i+1}, \omega_{i+2}, \dots, \omega_k, \beta) \\ + (-1)^{\varepsilon_{k-1}+1} (\alpha, \omega_1, \dots, \omega_{k-1}, \omega_k\beta),$$

where $\varepsilon_i = \deg \alpha + \deg \omega_1 + \dots + \deg \omega_i - i$. The differential δ has degree $+1$ with respect to the bar grading, while the differential d has degree $+1$ with respect to the tensor product grading. One may verify that $d\delta + \delta d = 0$, and we put $D \stackrel{\text{def}}{=} d + \delta$ to be the total differential. With the given bigrading, we get a double complex with the two differentials d and δ , which gives rise to the Eilenberg-Moore spectral sequence.

Chen's normalized version of this bar complex is the following. If $f \in A^0(B)$, let $S_i(f)$ be the operator on $B(A^\bullet(X), A^\bullet(B), A^\bullet(E))$ defined by

$$(51) \quad S_i(f)(\alpha, \omega_1, \dots, \omega_k, \beta) = (\alpha, \omega_1, \dots, \omega_{i-1}, f, \omega_i, \dots, \omega_k, \beta)$$

for $1 \leq i \leq k+1$. Let W be the subspace of $B(A^\bullet(X), A^\bullet(B), A^\bullet(E))$ generated by the images of $S_i(f)$ and $DS_i(f) - S_i(f)D$. Then define

$$(52) \quad \bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E)) \stackrel{\text{def}}{=} B(A^\bullet(X), A^\bullet(B), A^\bullet(E))/W.$$

Then W is closed under D , and when $H^0(B) = k$ (B is connected), then W is acyclic, so that $\bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E))$ is quasi-isomorphic to $B(A^\bullet(X), A^\bullet(B), A^\bullet(E))$. Notice that in the normalized bar complex there are no elements of negative degree, and with our assumption that B is simply-connected, we are assured convergence of the associated Eilenberg-Moore spectral sequence. The map θ mentioned above is induced by the map

$$(53) \quad \theta : B(A^\bullet(X), A^\bullet(B), A^\bullet(E)) \rightarrow A^\bullet(E_f),$$

which sends all tensor products to zero except for $A^\bullet(X) \otimes_k A^\bullet(E)$, where the map is $(\alpha, \beta) \mapsto \tilde{p}^* \alpha \wedge \tilde{f}^* \beta$. Note that $\theta(W) = 0$, so that we get an induced map

$$(54) \quad \theta : \bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E)) \rightarrow A^\bullet(E_f).$$

The normalized bar complex may also be augmented. The augmentation, ε , maps all elements of positive total degree to zero. The elements of degree zero have the form (f, g) , where $f \in A^0(X)$ and $g \in A^0(E)$. Then we define $\varepsilon(f, g) = \varepsilon_X(f)\varepsilon_E(g) = f(x_0)g(e_0)$, where x_0 and e_0 are chosen base points in X and E , respectively, and $\varepsilon_X, \varepsilon_E$ are the augmentations of $A^\bullet(X), A^\bullet(E)$, respectively. If we choose base points so that the pull-back diagram above preserves all base points, then θ is an augmentation-preserving map.

The bar complex has a natural coalgebra structure. Since we are inputting k CDGA's to the bar complex, we also obtain a structure of k CDGA on the bar complex via the shuffle product.

More specifically, if (a_1, \dots, a_p) and (b_1, \dots, b_q) are two ordered sets, then a *shuffle* σ of (a_1, \dots, a_p) with (b_1, \dots, b_q) is a permutation of the ordered set $(a_1, \dots, a_p, b_1, \dots, b_q)$ that preserves the order of the a_i 's as well as the order of the b_j 's. That is, we demand that if $i < j$, then $\sigma(a_i) < \sigma(a_j)$ and $\sigma(b_i) < \sigma(b_j)$.

We obtain a product on $B(A^\bullet(X), A^\bullet(B), A^\bullet(E))$ by first taking the normal tensor product on the $A^\bullet(X) \otimes A^\bullet(E)$ factors, then taking the tensor product of this product with the shuffle product on the $A^\bullet(B)^{\otimes i}$ factors. As usual, we introduce a sign $(-1)^{\deg(\alpha)\deg(\beta)}$ whenever α is moved past β . One checks that this product induces a product on Chen's normalized complex, $\bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E))$, as well. Thus we arrive at the following lemma, whose proof is left to the reader. For more details, see [16].

Lemma A.1. *Assume that we have the pull-back diagram (46), where p is a fibration and B is simply-connected. Then the normalized bar complex*

$$\bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E))$$

is a k CDGA. Moreover,

$$\theta : \bar{B}(A^\bullet(X), A^\bullet(B), A^\bullet(E)) \rightarrow A^\bullet(E_f)$$

is a quasi-isomorphism of k CDGA's.

Remark A.2. We note that Chen's normalization is functorial. That is, if we have a commutative diagram of k CDGA's

$$(55) \quad \begin{array}{ccccc} A_2 & \longleftarrow & B_2 & \longrightarrow & C_2 \\ \uparrow & & \uparrow & & \uparrow \\ A_1 & \longleftarrow & B_1 & \longrightarrow & C_1 \end{array}$$

then we get a map of k CDGA's $\bar{B}(A_1, B_1, C_1) \rightarrow \bar{B}(A_2, B_2, C_2)$.

We may formulate an equivariant analog of the bar complex. Let us consider again the pull-back diagram (46). If we suppose further that X, B , and E are G -spaces, and that f and p are equivariant maps, then we obtain a pull-back diagram

$$(56) \quad \begin{array}{ccc} EG \times_G E_f & \xrightarrow{\tilde{f}} & EG \times_G E \\ \tilde{p} \downarrow & & p \downarrow \\ EG \times_G X & \xrightarrow{f} & EG \times_G B \end{array}$$

Note that we are assuming B to be simply-connected, which in turn implies that $EG \times_G B$ is simply-connected as well. We may apply Lemma A.1 to the diagram (56). However, the bar complex $\bar{B}(A_G^\bullet(X), A_G^\bullet(B), A_G^\bullet(E))$ has the extra structure of an R CDGA^o or R CDGA, depending on fixed points. We may give it an R -algebra structure via the R -algebra structure on the $A_G^\bullet(X)$ factor, and we define the augmentation as above, assuming that we can choose our base points as described before to be actually fixed points of the group action. We arrive at the following.

Lemma A.3. *Assume that in the pull-back diagram (46), we have that X, B , and E are all G -spaces with f and p equivariant maps. Then the normalized bar complex*

$$\bar{B}(A_G^\bullet(X), A_G^\bullet(B), A_G^\bullet(E))$$

is an $RCDGA^\circ$. Moreover,

$$\theta : \bar{B}(A_G^\bullet(X), A_G^\bullet(B), A_G^\bullet(E)) \rightarrow A_G^\bullet(E_f)$$

is a quasi-isomorphism of $RCDGA^\circ$'s. If we assume further that all fixed-point sets are non-empty, and the diagram (46) preserves base points chosen from the various fixed-point sets, then the normalized bar complex is an $RCDGA$, and θ is a quasi-isomorphism of $RCDGA$'s.

In this equivariant case, we may further simplify the bar complex, following an idea of [10]. Let us consider the bar complex over R :

$$(57) \quad B_R(A_G^\bullet(X), A_G^\bullet(B), A_G^\bullet(E)) = \bigoplus_{i=0}^{\infty} A_G^\bullet(X) \otimes_R (sA_G^\bullet(B))^{\otimes i} \otimes_R A_G^\bullet(E),$$

where all the tensor products are over R .

Lemma A.4. Suppose that A, B , and C are $RCDGA$'s and we have morphisms of $RCDGA$'s $A \leftarrow B \rightarrow C$, where $R = H^\bullet(BG)$ for G a compact, connected Lie group. (We use this sequence to define a (differential graded) B -module structure on A and C .) Suppose, further, either that for each $r \in R$, r is not a zero-divisor in A , or that this condition holds for C . Then the natural projection

$$(58) \quad B_k(A, B, C) \rightarrow B_R(A, B, C)$$

is a quasi-isomorphism of $RCDGA$'s.

Proof. We have that $B_R(A, B, C) = B_k(A, B, C)/V$, where V is the sub-complex generated by all elements of the form

$$(59) \quad (a, b_1, \dots, rb_i, \dots, b_k, c) - (a, b_1, \dots, rb_{i+1}, \dots, b_k, c),$$

where $r \in R$, $a \in A$, $b_j \in B$, and $c \in C$. It is due to the fact that all elements of R have even degree that V is closed under the differential $D = d + \delta$. We claim that V is, in fact, acyclic. To see this, consider the map $s : V^i \rightarrow V^{i-1}$ defined by

$$(60) \quad s\{(a, b_1, \dots, rb_i, \dots, b_k, c) - (a, b_1, \dots, rb_{i+1}, \dots, b_k, c)\} \\ = (-1)^{\varepsilon_i} \{(a, b_1, \dots, rb_i, 1, b_{i+1}, \dots, b_k, c) - (a, b_1, \dots, b_i, r, b_{i+1}, \dots, b_k, c)\},$$

where $\varepsilon_i = \deg a + \deg \omega_1 + \dots + \deg \omega_i - i$. It is straightforward but tedious to check that $ds + sd = 0$, and that $\delta s + s\delta = id$, so that $Ds + sD = id$, and consequently V is acyclic. Moreover, it is easy to check that V is an ideal, so that the product on the bar complex induces a product on the bar complex over R . \square

Remark A.5. Lemma A.4 is valid using the normalized bar complex.

Corollary A.6. In the situation of Lemma A.3,

$$(61) \quad \theta : \bar{B}_R(A_G^\bullet(X), A_G^\bullet(B), A_G^\bullet(E)) \rightarrow A_G^\bullet(E_f)$$

is a quasi-isomorphism of $RCDGA^\circ$'s ($RCDGA$'s).

REFERENCES

1. C. Allday, *Rational homotopy and torus actions*, Houston Journal of Math. **5** (1979), 1–19. MR **80m**:57033
2. ———, *Invariant Sullivan-de Rham forms on cyclic sets*, J. London Math. Soc. **57** (1998), 478–490. MR **99g**:55015
3. C. Allday and V. Puppe, *Cohomological methods in transformation groups*, Cambridge University Press, 1993. MR **94g**:55009
4. A. K. Bousfield and V.K.A.M. Gugenheim, *On PL de Rham theory and rational homotopy type*, Memoirs of the American Mathematical Society, no. 179, 1976. MR **54**:13906
5. K. T. Chen, *Reduced bar constructions on de Rham complexes*, Algebra, Topology, and Category Theory, pp. 19–32, Academic Press, New York, 1976. MR **54**:1272
6. P. Deligne, P. Griffiths, J. Morgan, and D. Sullivan, *Real homotopy theory of Kähler manifolds*, Invent. Math. **29** (1975), 245–275. MR **52**:3584
7. W. G. Dwyer and J. Spalinski, *Homotopy theories and model categories*, Handbook of Algebraic Topology (I.M. James, ed.), Elsevier Science B.V., 1995, pp. 73–126. MR **96h**:55014
8. S. Eilenberg and J. C. Moore, *Homology and fibrations I, Coalgebras, cotensor product and its derived functors*, Comment. Math. Helvetica **40** (1966), 199–236. MR **34**:3579
9. A. Borel, *Seminar on transformation groups*, Ann. of Math. Studies, vol. 46, Princeton University Press, 1960. MR **22**:7129
10. E. Getzler, J. D. S. Jones, and S. Petrack, *Differential forms on loop spaces and the cyclic bar complex*, Topology **30** (1991), 339–371. MR **92i**:58179
11. M. Goresky, R. Kottwitz, and R. MacPherson, *Equivariant cohomology, Koszul duality, and the localization theorem*, Invent. Math. **131** (1998), 25–83. MR **99c**:55009
12. P.-P. Grivel, *Formes différentielles et suites spectrales*, Ann. Inst. Fourier **29** (1979), 17–37. MR **81b**:55041
13. V. Guillemin and S. Sternberg, *Supersymmetry and equivariant de Rham theory*, Springer-Verlag, New York, 1999. MR **2001i**:53140
14. S. Halperin, *Lectures on minimal models*, Mém. Soc. Math. France, vols. 9–10, 1983. MR **85i**:55009
15. S. Halperin and J. Stasheff, *Obstructions to homotopy equivalences*, Advances in Mathematics **32** (1979), 233–279. MR **80j**:55016
16. S. Lillywhite, *Bar complexes and formality of pull-backs*, Preprint.
17. ———, *The topology of symplectic quotients of loop spaces*, Ph.D. thesis, The University of Maryland, College Park, MD., 1998.
18. ———, *The topology of the moduli space of arc-length parametrised closed curves in Euclidean space*, Topology **39** (2000), 487–494. MR **2001b**:55028
19. G. Lupton, *Variations on a conjecture of Halperin*, Homotopy and Geometry, vol. 45, Banach Center Publications, 1998, pp. 115–135. MR **99m**:55014
20. J. McCleary, *User's guide to spectral sequences*, Publish or Perish, Inc., Wilmington, DE, 1985. MR **87f**:55014
21. T. J. Miller, *On the formality of $(k-1)$ -connected compact manifolds of dimension less than or equal to $4k-2$* , Illinois Journal of Math. **23** (1979), 253–258. MR **80j**:55017
22. D. G. Quillen, *Homotopical algebra*, Lecture Notes in Mathematics, vol. 43, Springer-Verlag, 1967. MR **36**:6480
23. ———, *Rational homotopy theory*, Ann. of Math. **90** (1969), 205–295. MR **41**:2678
24. L. Smith, *Homological algebra and the Eilenberg-Moore spectral sequence*, Trans. Amer. Math. Soc. **129** (1967), 58–93. MR **35**:7337
25. C. Teleman, *The quantization conjecture revisited*, Ann. of Math **152** (2000), 1–43. MR **2002d**:14073

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, 100 ST. GEORGE ST., TORONTO, ONTARIO, CANADA M5S 3G3

E-mail address: sml@math.toronto.edu

LARGE RECTANGULAR SEMIGROUPS IN STONE-ČECH COMPACTIFICATIONS

NEIL HINDMAN, DONA STRAUSS, AND YEVHEN ZELENYUK

ABSTRACT. We show that large rectangular semigroups can be found in certain Stone-Čech compactifications. In particular, there are copies of the $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$ rectangular semigroup in the smallest ideal of $(\beta\mathbb{N}, +)$, and so, a semigroup consisting of idempotents can be embedded in the smallest ideal of $(\beta\mathbb{N}, +)$ if and only if it is a subsemigroup of the $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$ rectangular semigroup. In fact, we show that for any ordinal λ with cardinality at most \mathfrak{c} , $\beta\mathbb{N}$ contains a semigroup of idempotents whose rectangular components are all copies of the $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$ rectangular semigroup and form a decreasing chain indexed by $\lambda + 1$, with the minimum component contained in the smallest ideal of $\beta\mathbb{N}$.

As a fortuitous corollary we obtain the fact that there are \leq_L -chains of idempotents of length \mathfrak{c} in $\beta\mathbb{N}$. We show also that there are copies of the direct product of the $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$ rectangular semigroup with the free group on $2^{\mathfrak{c}}$ generators contained in the smallest ideal of $\beta\mathbb{N}$.

1. INTRODUCTION

The Stone-Čech compactification of the integers $\beta\mathbb{N}$ has a semigroup structure which extends addition on \mathbb{N} and has significant applications in Ramsey Theory and topological dynamics. Some questions about the algebra of $\beta\mathbb{N}$, which sound deceptively simple, have been found to be extremely difficult. For example, it is not known whether $\beta\mathbb{N}$ contains any finite semigroups whose members are not all idempotent. Whether there were two idempotents in $\beta\mathbb{N}$ whose sum was an idempotent different from either remained an open question for several years. It was answered in the affirmative in [10], in which it was shown that a certain finite rectangular semigroup could be embedded in $\beta\mathbb{N}$. (A semigroup is *rectangular* if and only if it is isomorphic to the direct product of a left zero semigroup and a right zero semigroup. A *rectangular component* of a semigroup of idempotents is a maximal rectangular subsemigroup. As suggested by the name, distinct components are disjoint. The components are partially ordered by the relation $P \leq Q$ if and only if $PQ \subseteq P$, equivalently $QP \subseteq P$ [7, Theorem 1].) In this paper, we show that the rectangular semigroup $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$, with the first factor being left zero and the second right zero, can be embedded in $\beta\mathbb{N}$. Indeed, $\beta\mathbb{N}$ contains semigroups of idempotents which are the union of \mathfrak{c} rectangular components each isomorphic to $2^{\mathfrak{c}} \times 2^{\mathfrak{c}}$. We shall show also that if S is an infinite cancellative semigroup with cardinality κ , then $\beta S \setminus S$ contains a semigroup of idempotents which is the union

Received by the editors April 12, 2002 and, in revised form, November 14, 2002.

2000 *Mathematics Subject Classification.* Primary 20M10; Secondary 22A15, 54H13.

The first author acknowledges support received from the National Science Foundation (USA) via grant DMS-0070593.

of at least κ rectangular components, each isomorphic to $2^{2^\kappa} \times 2^{2^\kappa}$, with the first factor being left zero and the second right zero.

We first review terminology used in the topological theory of semigroups. Let S be a semigroup and a topological space. For each $s \in S$, we define mappings λ_s and ρ_s from S to itself by $\lambda_s(t) = st$ and $\rho_s(t) = ts$. S is said to be a *right topological* semigroup if ρ_s is continuous for every $s \in S$. In this case, the *topological center* $\Lambda(S)$ of S is defined by $\Lambda(S) = \{s \in S : \lambda_s \text{ is continuous}\}$. S is said to be a *semitopological* semigroup if ρ_s and λ_s are both continuous for every $s \in S$. It is said to be a *topological* semigroup if the semigroup operation is a continuous mapping from $S \times S$ to S .

If S is a discrete semigroup, we regard its Stone-Ćech compactification βS as the space of ultrafilters defined on S , with the topology defined by choosing the sets of the form $\overline{A} = \{p \in \beta S : A \in p\}$ as a base for the open sets. βS is then a compact Hausdorff space and $\overline{A} = \text{cl}_{\beta S}(A)$. We regard S as a subset of βS , by identifying each element of S with the principal ultrafilter that it defines. βS can be given a semigroup structure which extends the semigroup structure of S in such a way that βS is a compact right topological semigroup, with S contained in its topological center. If $A \subseteq S$, A^* will denote $\overline{A} \setminus A$.

We shall use basic algebraic properties that hold in all compact Hausdorff right topological semigroups. (We shall be assuming that all hypothesized topological spaces are Hausdorff.) A simple and important property is that every compact right topological semigroup T contains an idempotent. T has a smallest ideal $K(T)$, which is both the union of all the minimal left ideals and the union of all the minimal right ideals of T . Every right ideal of T contains a minimal right ideal, and every left ideal of T contains a minimal left ideal. If L is a minimal left ideal and R a minimal right ideal in T , then $RL = R \cap L$ is a group. So $R \cap L$ contains a unique idempotent. If $f : T \rightarrow T'$ is a homomorphism from T onto a compact right topological group T' , then $f[K(T)] = K(T')$. For each minimal right ideal R' of T' , there is a minimal right ideal R of T for which $f[R] = R'$. The corresponding statement holds for left ideals as well. There are three natural orderings of the idempotents of T defined by

$$\begin{aligned} e \leq_L f &\Leftrightarrow e = ef, \\ e \leq_R f &\Leftrightarrow e = fe, \text{ and} \\ e \leq f &\Leftrightarrow ef = fe = e. \end{aligned}$$

An idempotent e is minimal with respect to any or all of these orderings if and only if $e \in K(T)$. The reader is referred to [1], [6], or [9] for proofs of these statements.

When S is a discrete semigroup, the smallest ideal $K(\beta S)$ is of special importance for combinatorial applications, and in particular, the members of idempotents in $K(\beta S)$ have strong combinatorial properties. (See [6, Chapter 14].) Thus we are especially interested in those semigroups of idempotents that can be embedded in the smallest ideal of βS .

As we have already mentioned, a semigroup S is rectangular provided it is isomorphic to the direct product of a left zero semigroup with a right zero semigroup. This is equivalent to saying that it satisfies the identities $x^2 = x$ and $xyz = xz$. (The necessity is trivial. For the sufficiency, pick $x \in S$, note that Sx is a left zero semigroup, xS is a right zero semigroup, and the function $(a, b) \mapsto ab$ from $Sx \times xS$ to S is an isomorphism.) We observe that a rectangular semigroup S

satisfies $S = K(S) = LR \sim L \times R$, where L denotes any minimal left ideal and R any minimal right ideal in S .

If S is a semigroup, $E(S)$ will denote the set of idempotents in S .

If A is any set, $\mathcal{P}_f(A)$ will denote the set of finite nonempty subsets of A .

2. THE SEMIGROUPS \mathbb{H}_κ

The subsemigroup $\mathbb{H} = \bigcap_{n=1}^\infty \text{cl}_{\beta\mathbb{N}}(\mathbb{N}2^n)$ of $(\beta\mathbb{N}, +)$ holds all of the idempotents of $\beta\mathbb{N}$ and much of the known algebraic structure. (See [6, Section 6.1].) It occurs widely in the study of semigroups of the form βS . If S is an infinite discrete cancellative semigroup, every G_δ subset of S^* that contains an idempotent, contains copies of \mathbb{H} [6, Theorem 6.32]. \mathbb{H} also has the property that any compact right topological semigroup with countable dense topological center is the image of \mathbb{H} under a continuous homomorphism [6, Theorem 6.4]. In this section we introduce a semigroup \mathbb{H}_κ which satisfies a similar conclusion for an arbitrary infinite cardinal κ . As a consequence of the results of the next section we shall conclude that each \mathbb{H}_κ contains large rectangular subsemigroups.

Definition 2.1. Let κ be an infinite cardinal. Then $W_\kappa = \bigoplus_{\alpha < \kappa} \mathbb{Z}_2$. For $x \in W_\kappa$, $\text{supp}(x) = \{\alpha < \kappa : x_\alpha \neq 0\}$. For $\alpha < \kappa$, e_α is that member of W_κ such that $\text{supp}(e_\alpha) = \{\alpha\}$, and

$$\mathbb{H}_\kappa = \bigcap_{\alpha < \kappa} \text{cl}_{\beta W_\kappa} \{x \in W_\kappa \setminus \{0\} : \min \text{supp}(x) \geq \alpha\}.$$

The structure of \mathbb{H}_ω is that induced by an “oid” as introduced by John Pym [8]. When we say that two structures are “topologically and algebraically isomorphic”, we mean that there is one function between them that is both an isomorphism and a homeomorphism.

Theorem 2.2. *The compact right topological semigroups \mathbb{H} and \mathbb{H}_ω are topologically and algebraically isomorphic.*

Proof. [6, Theorem 6.15]. □

It is a fact [6, Lemma 6.8] that all of the idempotents of $\beta\mathbb{N}$ are in \mathbb{H} . Thus, by [6, Theorem 1.65], $K(\mathbb{H}) = K(\beta\mathbb{N}) \cap \mathbb{H}$.

Theorem 2.3. *Let S be a countably infinite discrete group. Then $\beta S \setminus S$ contains a topological and algebraic copy T of \mathbb{H} such that $K(T) = K(\beta S) \cap T$.*

Proof. Take any idempotent $p \in K(\beta S)$. By [6, Theorem 9.13], there is a left invariant zero-dimensional Hausdorff topology on S in which the ultrafilter p converges to 1. Then by [6, Theorem 7.24], with $X = G = V(a) = S$ for every $a \in G$, there is a topological and algebraic embedding $f : \mathbb{H} \rightarrow \beta S \setminus S$ such that $p \in f[\mathbb{H}]$. It remains to apply [6, Theorem 1.65]. □

A similar result applies to the semigroup (\mathbb{N}, \cdot) . Given $n \in \omega$ we define the *binary support* of n by $n = \sum_{t \in \text{supp}_2(n)} 2^t$ and $\text{supp}_2(0) = \emptyset$.

Theorem 2.4. *Let $S = (\mathbb{N}, \cdot)$. There is a topological and algebraic copy T of \mathbb{H} contained in $\beta S \setminus S$ which contains all of the idempotents of $\beta S \setminus S$. In particular, $K(T) = T \cap K(\beta S)$.*

Proof. Let $\langle p_i \rangle_{i=1}^\infty$ be the sequence of primes. Then $(\bigoplus_{i=1}^\infty \omega, +)$ is isomorphic to (\mathbb{N}, \cdot) via the map $x \mapsto \prod_{i=1}^\infty p_i^{x_i}$; so we shall take S to be $\bigoplus_{i=1}^\infty \omega$. For each $i \in \mathbb{N}$, let $\pi_i : S \rightarrow \omega$ be the projection to the i^{th} factor and let $\tilde{\pi}_i : \beta S \rightarrow \beta \omega$ be its continuous extension.

Let $\{X_i : i \in \mathbb{N}\}$ be a partition of ω into infinite sets and for each $i \in \mathbb{N}$, let $\phi_i : X_i \rightarrow \omega$ be a bijection. We define $\theta : \omega \rightarrow S$ by agreeing that for each $i \in \mathbb{N}$ and each $n \in \omega$,

$$\pi_i(\theta(n)) = \sum_{j \in \text{supp}_2(n) \cap X_i} 2^{\phi_i(j)},$$

where $\sum_{j \in \emptyset} 2^{\phi_i(j)} = 0$. We note that θ is a bijection. (If $j \in \text{supp}_2(n) \setminus \text{supp}_2(m)$, then for some i , $j \in X_i$ and so $\phi_i(j) \in \text{supp}_2(\pi_i(n)) \setminus \text{supp}_2(\pi_i(m))$. Also, given $x \in S$, for each $i \in \mathbb{N}$ let $Y_i = \phi_i^{-1}[\text{supp}_2(\pi_i(x))]$, let $Z = \bigcup_{i \in \mathbb{N}} Y_i$, and let $n = \sum_{t \in Z} 2^t$. Then $\theta(n) = x$.) Consequently by [6, Exercise 3.4.1] the continuous extension $\tilde{\theta} : \beta \omega \rightarrow \beta S$ of θ is a bijection. Since $\theta(n + m) = \theta(n) + \theta(m)$ if $\text{supp}_2(n) \cap \text{supp}_2(m) = \emptyset$, $\tilde{\theta}$ is a homomorphism on \mathbb{H} by [6, Lemma 6.3].

To complete the proof, let p be an idempotent in $\beta S \setminus S$. Since

$$\tilde{\theta}[\mathbb{H}] = \tilde{\theta}[\bigcap_{n=1}^\infty \overline{\omega 2^n} \setminus \{0\}] = (\bigcap_{n=1}^\infty \overline{\theta[\omega 2^n]} \setminus \{0\}),$$

it suffices to show that for all $n \in \mathbb{N}$, $\theta[\omega 2^n] \in p$. So let $n \in \mathbb{N}$ and suppose that $\theta[\omega 2^n] \notin p$. Pick $t \in \{1, 2, \dots, 2^n - 1\}$ such that $\theta[\omega 2^n + t] \in p$, pick $j \in \text{supp}_2(t)$, and pick i such that $j \in X_i$. Now $\tilde{\pi}_i(p)$ is an idempotent; so either $\tilde{\pi}_i(p) = 0$ or $\tilde{\pi}_i(p) \in \beta \mathbb{N} \setminus \mathbb{N}$. Thus by [6, Lemma 6.6] $\omega 2^{\phi_i(j)+1} \in p$. Pick $x \in \omega 2^{\phi_i(j)+1} \cap \pi_i[\theta[\omega 2^n + t]]$ and pick $k \in \omega 2^n + t$ such that $x = \pi_i(\theta(k))$. Then $j \in \text{supp}_2(k) \cap X_i$; so $\phi_i(j) \in \text{supp}_2(x)$, contradicting the fact that $x \in \omega 2^{\phi_i(j)+1}$. \square

Observe that if $\kappa > \omega$, then $\mathbb{H}_\kappa \cap K(\beta W_\kappa) = \emptyset$. To see this, one lets $p \in \mathbb{H}_\kappa$ and

$$q \in \bigcap_{k < \omega} \text{cl} \{x \in W_\kappa \setminus \{0\} : \min \text{supp}(x) \geq k \text{ and } \max \text{supp}(x) < \omega\}.$$

Then $p \notin \beta W_\kappa + q + p$ and so [6, Theorem 4.39] applies.

Theorem 2.5. *Let κ be an infinite cardinal and let T be a compact right topological semigroup. Assume that there is a set $A \subseteq \Lambda(T)$ such that $|A| \leq \kappa$ and A is dense in T . Then there is a continuous surjective homomorphism $f : \mathbb{H}_\kappa \rightarrow T$.*

Proof. Enumerate A as $\{t_\alpha : \alpha < \kappa\}$, with repetition if $|A| < \kappa$. Let $\{I_\gamma : \gamma < \kappa\}$ be a partition of κ into subsets of size κ . Define $h : W_\kappa \rightarrow T$ by first agreeing that for each $\alpha < \kappa$, $h(e_\alpha) = t_\gamma$, where $\alpha \in I_\gamma$. Then for $F \in \mathcal{P}_f(\kappa)$, define $h(\sum_{\alpha \in F} e_\alpha) = \prod_{\alpha \in F} h(e_\alpha)$, where the product is taken in increasing order of indices. Define $h(0)$ arbitrarily. Let $\tilde{h} : \beta W_\kappa \rightarrow T$ be the continuous extension of h and let f be the restriction of \tilde{h} to \mathbb{H}_κ .

To see that $h[\mathbb{H}_\kappa] = T$, it suffices to show that $A \subseteq h[\mathbb{H}_\kappa]$. Given $\gamma < \kappa$, we have that $|I_\gamma| = \kappa$. Pick a κ -uniform ultrafilter p on $\{e_\alpha : \alpha \in I_\gamma\}$. Then $p \in \mathbb{H}_\kappa$ and $h(p) = t_\gamma$ because f is constantly equal to t_γ on $\{e_\alpha : \alpha \in I_\gamma\}$.

To see that f is a homomorphism it suffices by [6, Theorem 4.21] to observe that whenever $x \in W_\kappa \setminus \{0\}$ and $y \in W_\kappa \setminus \{0\}$ with $\min \text{supp}(y) > \max \text{supp}(x)$, then $h(x + y) = h(x) \cdot h(y)$. \square

Definition 2.6. Let S be a semigroup, let κ be a cardinal, and let $\langle t_\lambda \rangle_{\lambda < \kappa}$ be a κ -sequence in S .

- (a) Given $F \in \mathcal{P}_f(\kappa)$, $\prod_{\lambda \in F} t_\lambda$ is the product in increasing order of indices.

- (b) If $D \subseteq \kappa$, then $FP(\langle t_\lambda \rangle_{\lambda \in D}) = \{\prod_{\lambda \in F} t_\lambda : F \in \mathcal{P}_f(D)\}$.
- (c) The sequence $\langle t_\lambda \rangle_{\lambda < \kappa}$ has *distinct finite products* if and only if whenever $F, G \in \mathcal{P}_f(\lambda)$ and $\prod_{\lambda \in F} t_\lambda = \prod_{\lambda \in G} t_\lambda$, one must have $F = G$.

Theorem 2.7. *Let S be an infinite cancellative discrete semigroup with cardinality κ . Then $\beta S \setminus S$ contains a topological and algebraic copy of \mathbb{H}_κ .*

Proof. By [6, Lemma 6.31], we may choose a κ -sequence $\langle t_\lambda \rangle_{\lambda < \kappa}$ in S with distinct finite products. Let $T = FP(\langle t_\lambda \rangle_{\lambda < \kappa})$. For each $\gamma < \kappa$, let $T_\gamma = FP(\langle t_\lambda \rangle_{\gamma < \lambda < \kappa})$. We put $\tilde{T} = \bigcap_{\gamma < \kappa} \text{cl}_{\beta S}(T_\gamma)$. By [6, Theorem 4.20], \tilde{T} is a subsemigroup of βS .

We define $\theta : T \rightarrow W_\kappa$ by $\theta(\prod_{\lambda \in F} t_\lambda) = \sum_{\lambda \in F} e_\lambda$. (Since the sequence $\langle t_\lambda \rangle_{\lambda < \kappa}$ has distinct finite products, the function θ is well defined.) Let $\tilde{\theta} : \text{cl}_{\beta S}(T) \rightarrow \beta W_\kappa$ denote the continuous extension of θ . By [6, Theorem 4.21], the restriction of $\tilde{\theta}$ to \tilde{T} is a homomorphism. Now $\tilde{\theta}$ is injective, by [6, Exercise 3.4.1]. Since for each $\gamma < \kappa$, $\tilde{\theta}[T_\gamma] = \text{cl}_{\beta W_\kappa}\{x \in W_\kappa : \min \text{supp}(x) > \gamma\}$, $\tilde{\theta}$ maps \tilde{T} onto \mathbb{H}_κ . Thus $\tilde{\theta}$ determines an isomorphism from \tilde{T} onto \mathbb{H}_κ . \square

3. CHAINS OF RECTANGULAR SEMIGROUPS IN \mathbb{H}_κ

Let κ be an infinite cardinal and let V_κ denote the rectangular semigroup $\kappa \times \kappa$, with the first factor being left zero and the second right zero. We show in this section (in Corollary 3.10) that for any infinite cardinal κ , algebraic copies of $V_{2^{2^\kappa}}$ can be found in $K(\mathbb{H}_\kappa)$. Indeed, if λ is any ordinal for which $|\lambda| \leq \kappa$, there is a decreasing chain $\langle D_p \rangle_{p \leq \lambda}$ of disjoint copies of $V_{2^{2^\kappa}}$ contained in \mathbb{H}_κ , with D_λ embedded in $K(\mathbb{H}_\kappa)$.

Notice that V_κ contains a copy of every rectangular semigroup of cardinality at most κ .

Definition 3.1. Let λ be any ordinal and let A be any nonempty set. Let 0 denote a selected element of A . For $p < \lambda$, let $C_p = A \times A \times \{p\}$ and let

$$C = C_{A,\lambda} = A \cup \bigcup_{p < \lambda} C_p = A \cup (A \times A \times \lambda).$$

The operation \cdot on C is defined as follows. Let $a, b, c, d \in A$ and let $p, q < \lambda$. Then

$$\begin{aligned} a \cdot b &= b, \\ a \cdot (b, c, p) &= (b, c, p), \\ (b, c, p) \cdot a &= (b, a, p), \\ (a, b, p) \cdot (c, d, q) &= (a, d, p \vee q), \end{aligned}$$

where $p \vee q$ is the maximum of p and q .

We leave to the reader the routine verification that the operation on $C_{A,\lambda}$ is associative. Notice that for any $p < \lambda$, C_p is a copy of $V_{|A|}$.

Definition 3.2. Let λ be an ordinal and let $p < \lambda$. We let $u_p = (0, 0, p)$, and for every $x = (a, b, p) \in C_p$, we let $x_\ell = (a, 0, 0)$ and $x_r = b$. For $x \in A$, we let $x_\ell = x_r = x$.

The following is simple, and its proof is like that of [6, Theorem 1.46].

Lemma 3.3. *Let S be a semigroup, let H be an ideal of S , let L be a minimal left ideal of H , let R be a minimal right ideal of H , and let $x \in S$. Then Lx is a minimal left ideal of H , xR is a minimal right ideal of H , $xL \subseteq L$, and $Rx \subseteq R$.*

Proof. It suffices to establish the assertions about Lx and xL . Now $Lx \subseteq Hx \subseteq H$ and $HLx \subseteq Lx$; so Lx is a left ideal of H . Let M be a left ideal of H with $M \subseteq Lx$. Let $J = \{y \in L : yx \in M\}$. Given $y \in J$ and $z \in H$, we have $zy \in L$ and $zyx \in M$; so $zy \in J$. Thus $J = L$ and so $M = Lx$.

Next, given $y \in L = Hy$, so pick $z \in H$ such that $y = zy$. Then $xy = xzy \in Hy = L$. \square

Lemma 3.4. *Let A be a nonempty set with distinguished element 0, and let $C = C_{A,1}$. Let T be a right topological semigroup, and let $f : T \rightarrow C$ be a surjective homomorphism for which $f^{-1}[A]$ and $f^{-1}[C_0]$ are compact. Then there is a homomorphism $g : C \rightarrow T$ such that $f \circ g$ is the identity on C and $g[C_0] \subseteq K(f^{-1}[C_0])$. If T is compact, then $g[C_0] \subseteq K(T)$.*

Proof. We first define g on A . We have that $f^{-1}[A]$ is a compact semigroup. Choose a minimal right ideal N of $f^{-1}[A]$. For each $a \in A$, $f^{-1}[\{a\}]$ is a left ideal of $f^{-1}[A]$. So choose a minimal left ideal S_a of $f^{-1}[A]$ with $S_a \subseteq f^{-1}[\{a\}]$, and let $g(a)$ be the identity of the group $N \cap S_a$. Then immediately $f(g(a)) = a$. Also, given $a, b \in A$ we have that $g(a)$ and $g(b)$ are idempotents in N ; so $g(a)g(b) = g(b) = g(ab)$.

Let $B = \{(a, 0, 0) : a \in A\}$. Then B is a left ideal of C_0 ; so $f^{-1}[B]$ is a left ideal of $f^{-1}[C_0]$ which therefore contains a minimal left ideal L of $f^{-1}[C_0]$. For each $a \in A$ let $F_a = \{(a, b, 0) : b \in A\}$. Then F_a is a right ideal of C_0 . So pick a minimal right ideal R_a of $f^{-1}[C_0]$ with $R_a \subseteq f^{-1}[F_a]$. By Lemma 3.3, since $f^{-1}[C_0]$ is an ideal of $f^{-1}[A \cup C_0]$, we have that $g(0) \cdot R_a$ is a minimal right ideal of $f^{-1}[C_0]$ and $L \cdot g(a)$ is a minimal left ideal of $f^{-1}[C_0]$. For $a, b \in A$, let $g(a, b, 0)$ be the identity of the group $g(0) \cdot R_a \cdot L \cdot g(b)$. Notice that if T is compact, then $K(T) \subseteq f^{-1}[C_0]$ so that $K(f^{-1}[C_0]) \subseteq K(T)$ and thus $g(a, b, 0) \in K(T)$. Also $g(a, b, 0) = g(0) \cdot x \cdot y \cdot g(b)$ for some $x \in R_a$ and some $y \in L$. So $f(g(a, b, 0)) = 0 \cdot f(x) \cdot f(y) \cdot b = (a, b, 0)$.

To conclude the proof we need to show that g is a homomorphism. First we let $a, b, c \in A$ and show that $g(a) \cdot g(b, c, 0) = g(b, c, 0)$ and $g(b, c, 0) \cdot a = g(b, a, 0)$. Pick $x \in R_b \cdot L$ such that $g(b, c, 0) = g(0) \cdot x \cdot g(c)$. Then

$$\begin{aligned} g(a) \cdot g(b, c, 0) &= g(a) \cdot g(0) \cdot x \cdot g(c) \\ &= g(0) \cdot x \cdot g(c) \\ &= g(b, c, 0). \end{aligned}$$

So the first claim holds directly. Multiplying on the left by $g(b, c, 0)$ and on the right by $g(a)$ one sees that $g(b, c, 0) \cdot g(a)$ is idempotent. Since $g(b, c, 0) \cdot g(a) \in g(0) \cdot R_b \cdot L \cdot g(c) \cdot g(a) = g(0) \cdot R_b \cdot L \cdot g(a)$, we must have that $g(b, c, 0) \cdot g(a)$ is the identity of $g(0) \cdot R_b \cdot L \cdot g(a)$, namely $g(b, a, 0)$.

Finally, let $a, b, c, d \in A$. Then $g(a, b, 0) \cdot g(c, d, 0) \in g(0) \cdot R_a \cdot L \cdot g(b) \cdot g(0) \cdot R_c \cdot L \cdot g(d) \subseteq g(0) \cdot R_a \cdot L \cdot g(d)$. So it suffices to show that $g(a, b, 0) \cdot g(c, d, 0)$ is idempotent. These elements satisfy $g(a, 0, 0) \in L \cdot g(0)$ and $g(c, 0, 0) \in L \cdot g(0)$. So as idempotents in the same minimal left ideal of $f^{-1}[C_0]$, we have that $g(a, 0, 0) \cdot g(c, 0, 0) = g(a, 0, 0)$. Recall that we have shown that for any $x, y \in A$, $g(0) \cdot g(x, y, 0) = g(x, y, 0)$ and $g(x, y, 0) \cdot g(0) = g(x, 0, 0)$. Thus we have

$$\begin{aligned} g(a, b, 0) \cdot g(c, d, 0) \cdot g(a, b, 0) &= g(a, b, 0) \cdot g(0) \cdot g(c, d, 0) \cdot g(0) \cdot g(a, b, 0) \\ &= g(a, 0, 0) \cdot g(c, 0, 0) \cdot g(a, b, 0) \\ &= g(a, 0, 0) \cdot g(a, b, 0) \\ &= g(a, b, 0) \cdot g(0) \cdot g(a, b, 0) \\ &= g(a, b, 0) \cdot g(a, b, 0) = g(a, b, 0). \end{aligned}$$

Multiplying on the right by $g(c, d, 0)$ we have that $g(a, b, 0) \cdot g(c, d, 0)$ is idempotent. \square

We now consider the situation in which $\lambda > 1$. For $\lambda > \omega$ we do not necessarily get that g is a homomorphism, but we come close.

Theorem 3.5. *Let A be a nonempty set with distinguished element 0, let λ be an ordinal, and let $C = C_{A,\lambda}$. Let T be a right topological semigroup, and $f : T \rightarrow C$ be a surjective homomorphism such that $f^{-1}[A]$ is compact and $f^{-1}[C_p]$ is compact for every $p < \lambda$. Then there is a function $g : C \rightarrow T$ such that $f \circ g$ is the identity and g has the following properties:*

- (i) *If $q \leq p < \lambda$, $x \in C_p$, and $y \in A \cup C_q$, then $g(xy) = g(x) \cdot g(y)$ and $g(y) \cdot g(x)$ is an idempotent in the same minimal left ideal of $f^{-1}[C_p]$ as $g(yx)$.*
- (ii) *If $p < \lambda$, $x \in C_p$, and $y \in A \cup C_0$, then $g(y) \cdot g(x) = g(yx)$.*
- (iii) *If $q \leq p < \lambda$, $n \in \omega$, $p = q + n$, $y \in C_q$, and $x \in C_p$, then $g(y) \cdot g(x) = g(yx)$.*
- (iv) *If $p < \lambda$, then $g[C_p] \subseteq K(f^{-1}[C_p])$.*
- (v) *If T is compact and λ is a successor, then $g[C_{\lambda-1}] \subseteq K(T)$.*

The semigroup T contains a semigroup $D = \bigcup_{p < \lambda} D_p$ of idempotents where each D_p is a rectangular component of D with $g[C_p] \subseteq D_p$ and the sequence $\langle D_p \rangle_{p < \lambda}$ is decreasing in the ordering of components, so that for each $p < \lambda$, $D_p = K(\bigcup_{q \leq p} D_q)$. If $|A| \geq |\lambda| \geq \omega$, then for each $p < \lambda$, $|D_p|$ is isomorphic to $V_{|A|}$.

Proof. For $p < \lambda$ we define g on $A \cup \bigcup_{q \leq p} C_q$ by induction on p so that g satisfies conclusions (i), (ii), (iii), and (iv). By Lemma 3.4 we may define g on $A \cup C_0$ so that g satisfies (iv) and is a homomorphism and therefore satisfies (i), (ii), and (iii). Now let $0 < p < \lambda$ and assume that g has been defined on $A \cup \bigcup_{q < p} C_q$.

We show first that we may choose a minimal left ideal L of $f^{-1}[C_p]$ such that $L \subseteq \bigcap_{q < p} f^{-1}[C_p] \cdot g(u_q)$ and $f[L] = C_p \cdot u_p$. A simple computation establishes that for each $q < p$, $f^{-1}[C_p] \cdot g(u_q)$ is a compact left ideal of $f^{-1}[C_p]$. Also, if $r < q < p$, then $f^{-1}[C_p] \cdot g(u_q) \subseteq f^{-1}[C_p] \cdot g(u_r)$. To see this, let $x \in f^{-1}[C_p]$. Then

$$\begin{aligned} x \cdot g(u_q) &= x \cdot g(u_q \cdot u_r) \\ &= x \cdot g(u_q) \cdot g(u_r) \quad \text{by (i)} \\ &\in f^{-1}[C_p] \cdot g(u_r). \end{aligned}$$

Consequently, $\bigcap_{q < p} f^{-1}[C_p] \cdot g(u_q)$ is a left ideal of $f^{-1}[C_p]$ and thus contains a minimal left ideal L of $f^{-1}[C_p]$. Then $f[L]$ is a minimal left ideal of C_p .

Now given $x \in L$, one has $x \in f^{-1}[C_p] \cdot g(u_0)$. So for some $a \in A$, $f(x) = (a, 0, p) \in C_p \cdot u_p$. Thus $f[L] \subseteq C_p \cdot u_p$. Since $C_p \cdot u_p$ is a minimal left ideal of C_p , we have $f[L] = C_p \cdot u_p$ as claimed.

If p is a successor ordinal, observe that $g(u_{p-1}) \cdot f^{-1}[C_p]$ is a right ideal of $f^{-1}[C_p]$ and pick a minimal right ideal R of $f^{-1}[C_p]$ with $R \subseteq g(u_{p-1}) \cdot f^{-1}[C_p]$. Then $f[R]$ is a minimal right ideal of C_p and $f[R] \subseteq u_{p-1} \cdot C_p$; so $f[R] = u_{p-1} \cdot C_p$.

If p is a limit ordinal, note that $f^{-1}[u_p \cdot C_p]$ is a right ideal of $f^{-1}[C_p]$. So pick a minimal right ideal R of $f^{-1}[C_p]$ with $R \subseteq f^{-1}[u_p \cdot C_p]$. Then $f[R] = u_p \cdot C_p$.

Now $f^{-1}[C_p]$ is an ideal of $f^{-1}[A \cup C_0 \cup C_p]$. So by Lemma 3.3, for any $x \in C_p$, $g(x_\ell) \cdot R$ is a minimal right ideal of $f^{-1}[C_p]$ and $L \cdot g(x_r)$ is a minimal left ideal of $f^{-1}[C_p]$. Therefore $(g(x_\ell) \cdot R) \cap (L \cdot g(x_r)) = g(x_\ell) \cdot R \cdot L \cdot g(x_r)$ is a group. Let $g(x)$ be the identity of $g(x_\ell) \cdot R \cdot L \cdot g(x_r)$. Notice that (iv) is satisfied.

To verify (i), let $q \leq p$, let $x \in C_p$, and let $y \in A \cup C_q$. To see that $g(xy) = g(x) \cdot g(y)$, we show that $g(x) \cdot g(y)$ is an idempotent in the same group as $g(xy)$. Since $(xy)_\ell = x_\ell$ and by Lemma 3.3, $R \cdot g(y) \subseteq R$, we have that $g(xy) \in g((xy)_\ell) \cdot R = g(x_\ell) \cdot R$ and $g(x) \cdot g(y) \in g(x_\ell) \cdot R \cdot g(y) \subseteq g(x_\ell) \cdot R$.

Also, $(xy)_r = y_r$ and so $g(xy) \in L \cdot g((xy)_r) = L \cdot g(y_r)$. To see that $g(x) \cdot g(y) \in L \cdot g(y_r)$, we consider two cases. If $y \in C_p$, then $g(x) \cdot g(y) \in g(x) \cdot L \cdot g(y_r) \subseteq L \cdot g(y_r)$. Now assume that $q < p$ (and $y \in A \cup C_q$). Note that $L \subseteq f^{-1}[c_p] \cdot g(u_q)$; so $g(u_q)$ is a right identity for L and thus $L = L \cdot g(u_q)$. Also, a simple computation establishes that $u_q x_r y = u_q y_r$. Therefore, using the fact that (i) holds at q , $g(u_q) \cdot g(x_r) \cdot g(y) = g(u_q x_r) \cdot g(y) = g(u_q x_r y) = g(u_q y_r) = g(u_q) \cdot g(y_r)$ and thus $g(x) \cdot g(y) \in L \cdot g(x_r) \cdot g(y) = L \cdot g(u_q) \cdot g(x_r) \cdot g(y) = L \cdot g(u_q) \cdot g(y_r) = L \cdot g(y_r)$.

Consequently, we have in any event that $g(xy)$ and $g(x) \cdot g(y)$ are members of the group $g(x_\ell) \cdot R \cdot L \cdot g(y_r)$. We show that they are equal by showing that $g(x) \cdot g(y)$ is idempotent.

Since $g(x_r) \cdot g(0) = g(x_r 0) = g(0) = g(y_r 0) = g(y_r) \cdot g(0)$ we have that $g(x) \cdot g(0) \in L \cdot g(x_r) \cdot g(0) = L \cdot g(0)$ and because $g(x) \cdot g(y) \in L \cdot g(y_r)$ we have that $g(x) \cdot g(y) \cdot g(0) \in L \cdot g(y_r) \cdot g(0) = L \cdot g(0)$.

Now $g(x) = g(x_\ell) \cdot z \cdot g(x_r)$ for some $z \in L \cdot R$. Also $g(0) \cdot g(x_\ell) = g(0 x_\ell) = g(x_\ell)$ and so $g(0) \cdot g(x) = g(x_\ell) \cdot z \cdot g(x_r) = g(x)$. If $y \in C_p$ we have similarly that $g(0) \cdot g(y) = g(y)$, while otherwise $g(0) \cdot g(y) = g(0 y) = g(y)$ by (ii) of the induction hypothesis. We have that $g(x) \cdot g(0) \cdot g(x) \cdot g(0) = g(x) \cdot g(x) \cdot g(0) = g(x) \cdot g(0)$. So $g(x) \cdot g(0)$ is an idempotent in $L \cdot g(0)$, which is a minimal left ideal of $f^{-1}[c_p]$ by Lemma 3.3. Therefore $g(x) \cdot g(0)$ is a right identity for $L \cdot g(0)$ and thus $g(x) \cdot g(y) \cdot g(0) \cdot g(x) \cdot g(0) = g(x) \cdot g(y) \cdot g(0)$. So

$$\begin{aligned} g(x) \cdot g(y) \cdot g(x) \cdot g(y) &= g(x) \cdot g(y) \cdot g(0) \cdot g(x) \cdot g(0) \cdot g(y) \\ &= g(x) \cdot g(y) \cdot g(0) \cdot g(y) \\ &= g(x) \cdot g(y) \cdot g(y) = g(x) \cdot g(y), \end{aligned}$$

as required.

By Lemma 3.3, $g(y) \cdot g(x_\ell) \cdot R \subseteq f^{-1}[c_p]$; so $g(y) \cdot g(x) \in g(y) \cdot g(x_\ell) \cdot R \cdot L \cdot g(x_r) \subseteq L \cdot g(x_r)$. Also, $g(yx) \in L \cdot g((yx)_r) = L \cdot g(x_r)$ and by Lemma 3.3, $L \cdot g(x_r)$ is a minimal left ideal of $f^{-1}[c_p]$. To see that $g(y) \cdot g(x)$ is idempotent, note that $xyx = x$. So $g(x) \cdot g(y) \cdot g(x) = g(xy) \cdot g(x) = g(xyx) = g(x)$ and thus $g(y) \cdot g(x) \cdot g(y) \cdot g(x) = g(y) \cdot g(x)$, as required. This completes the verification of (i).

To verify (ii), let $x \in C_p$ and let $y \in A \cup C_0$. Pick $z \in L \cdot R$ such that $g(x) = g(x_\ell) \cdot z \cdot g(x_r)$. Then $g(y) \cdot g(x_\ell) = g(yx_\ell)$. If $y \in A$, then $yx_\ell = x_\ell$ so that $g(y) \cdot g(x) = g(y) \cdot g(x_\ell) \cdot z \cdot g(x_r) = g(x_\ell) \cdot z \cdot g(x_r) = g(x) = g(yx)$. So assume that $y \in C_0$. Then $yx_\ell = (yx)_\ell$ and $(yx)_r = x_r$. So $g(y) \cdot g(x) = g(y) \cdot g(x_\ell) \cdot z \cdot g(x_r) = g(yx_\ell) \cdot z \cdot g(x_r) = g((yx)_\ell) \cdot z \cdot g((yx)_r)$. So to see that $g(y) \cdot g(x) = g(yx)$ it suffices to recall from (i) that $g(y) \cdot g(x)$ is idempotent.

To verify (iii), let $n \in \omega$ and let $q \leq p$ such that $p = q + n$. Let $x \in C_p$ and let $y \in C_q$. If $n = 0$, the conclusion follows from (i). So assume that $n > 0$ so that p is a successor ordinal and $p - 1 \geq q$. Now $(yx)_\ell = y_\ell$; so $g(yx) \in g((yx)_\ell) \cdot R = g(y_\ell) \cdot R$.

Recall that $R \subseteq g(u_{p-1}) \cdot f^{-1}[C_p]$ and consequently $R = g(u_{p-1}) \cdot R$. Thus,

$$\begin{aligned}
 g(y) \cdot g(x) &\in g(y) \cdot g(x_\ell) \cdot R \\
 &= g(y) \cdot g(x_\ell) \cdot g(u_{p-1}) \cdot R \\
 &= g(yx_\ell) \cdot g(u_{p-1}) \cdot R && \text{by (i) at } q \\
 &= g(yx_\ell u_{p-1}) \cdot R && \text{by (iii) at } p-1 \\
 &= g(y_\ell u_{p-1}) \cdot R \\
 &= g(y_\ell) \cdot g(u_{p-1}) \cdot R && \text{by (ii) at } p-1 \\
 &= g(y_\ell) \cdot R.
 \end{aligned}$$

Now $g(y) \cdot g(x)$ is an idempotent in the same minimal left ideal of $f^{-1}[C_p]$ by (i). Since $g(yx)$ and $g(y) \cdot g(x)$ are also in the same minimal right ideal $g(y_\ell) \cdot R$, they must be equal. This completes the induction step.

Next, we establish (v). So assume that T is compact and λ is a successor. Then $K(T) \subseteq f^{-1}[C_{\lambda-1}]$ and so $K(f^{-1}[C_{\lambda-1}]) \subseteq K(T)$. Since $g(x) \in K(f^{-1}[C_{\lambda-1}])$ for every $x \in C_{\lambda-1}$, (v) holds.

For each $p < \lambda$, let

$$D_p = \{\prod_{q \in F} g(x_q) : F \in \mathcal{P}_f(\lambda), p = \max F, \text{ and for each } q \in F, x_q \in C_q\},$$

where for each F , the product $\prod_{q \in F} g(x_q)$ is taken in increasing order of indices.

We show now by induction on $|F|$ that

$$\begin{aligned}
 (*) \quad &\text{if } p < \lambda, y \in C_p, F \in \mathcal{P}_f(\lambda), x_q \in C_q \text{ for each } q \in F, \\
 &\text{and } \max F \leq p, \text{ then } g(y) \cdot \prod_{q \in F} g(x_q) = g(y \cdot \prod_{q \in F} x_q).
 \end{aligned}$$

Let $r = \max F$. If $F = \{r\}$, then $g(y) \cdot g(x_r) = g(yx_r)$ by (i). So assume that $|F| > 1$ and let $G = F \setminus \{r\}$. Then

$$\begin{aligned}
 g(y) \cdot \prod_{q \in F} g(x_q) &= g(y) \cdot \prod_{q \in G} g(x_q) \cdot g(x_r) \\
 &= g(y \cdot \prod_{q \in G} x_q) \cdot g(x_r) && \text{by the induction hypothesis} \\
 &= g(y \cdot \prod_{q \in G} x_q \cdot x_r) && \text{by (i).}
 \end{aligned}$$

Now we show that each member of D is idempotent. So let $F \in \mathcal{P}_f(\lambda)$, let $p = \max F$, and for each $q \in F$, let $x_q \in C_q$. If $F = \{p\}$, then $g(x_p)$ is idempotent. So assume that $|F| > 1$ and let $G = F \setminus \{p\}$. Then

$$\begin{aligned}
 \prod_{q \in F} g(x_q) \cdot \prod_{q \in F} g(x_q) &= \prod_{q \in G} g(x_q) \cdot g(x_p) \cdot \prod_{q \in F} g(x_q) \\
 &= \prod_{q \in G} g(x_q) \cdot g(x_p \cdot \prod_{q \in F} x_q) \\
 &= \prod_{q \in G} g(x_q) \cdot g(x_p \cdot \prod_{q \in G} x_q \cdot x_p) \\
 &= \prod_{q \in G} g(x_q) \cdot g(x_p) \\
 &= \prod_{q \in F} g(x_q).
 \end{aligned}$$

Now let $r, p < \lambda$, let $a \in D_r$, and let $b \in D_p$. We claim that $ab \in D_{r \vee p}$. We have that $a = \prod_{q \in F} g(x_q)$ and $b = \prod_{q \in G} g(y_q)$ where $\max F = r$, $\max G = p$, each $x_q \in C_q$ and each $y_q \in C_q$. If $r < \min G$, then $ab = \prod_{q \in F} g(x_q) \cdot \prod_{q \in G} g(y_q) \in D_p$. If $r \geq p$, then $ab = \prod_{q \in F \setminus \{r\}} g(x_q) \cdot g(x_r \cdot \prod_{q \in G} y_q) \in D_r$ (where the $\prod_{q \in F \setminus \{r\}} g(x_q)$ term is simply omitted if $F = \{r\}$). So assume that $\min G \leq r < p$, let $H = \{q \in G : q \leq r\}$, and let $L = \{q \in G : q > r\}$. Then $ab = \prod_{q \in F \setminus \{r\}} g(x_q) \cdot g(x_r \cdot \prod_{q \in H} y_q \cdot \prod_{q \in L} g(y_q)) \in D_p$.

Thus D is a semigroup of idempotents and for each $p < \lambda$, $K(\bigcup_{q \leq p} D_q) \subseteq D_p$. Let $r \leq p$, let $a = \prod_{q \in F} g(x_q) \in D_p$, let $b = \prod_{q \in G} g(y_q) \in D_r$, and let $c =$

$\prod_{q \in H} g(z_q) \in D_p$. Then

$$\begin{aligned} abc &= \prod_{q \in F \setminus \{p\}} g(x_q) \cdot g(x_p \cdot \prod_{q \in G} y_q) \cdot \prod_{q \in H} g(z_q) \\ &= \prod_{q \in F \setminus \{p\}} g(x_q) \cdot g(x_p \cdot \prod_{q \in G} y_q \cdot \prod_{q \in H} z_q) \\ &= \prod_{q \in F \setminus \{p\}} g(x_q) \cdot g(x_p z_p) \end{aligned}$$

and

$$\begin{aligned} ac &= \prod_{q \in F \setminus \{p\}} g(x_q) \cdot g(x_p \cdot \prod_{q \in H} z_q) \\ &= \prod_{q \in F \setminus \{p\}} g(x_q) \cdot g(x_p z_p). \end{aligned}$$

So $abc = ac$ and so each D_p is a rectangular subsemigroup of D . To see that D_p is a rectangular component of D , suppose that $a \in D_p$ and $b \in D_q$, where $q < p$. Then $f(bab) \in C_p$ and $f(b) \in C_q$, and so $bab \neq b$. To show that $D_p \approx V_{|A|}$ if $|A| \geq |\lambda| \geq \omega$, we observe that C_p contains a left ideal L and a right ideal R , each with $|A|$ elements. If $a, b \in L$, then $ab = a$ and so $g(a)g(b) = g(a)$. Thus $g[L]$ is contained in the left ideal $D_p g(b)$ of D_p . Similarly, $g[R]$ is contained in a right ideal of D_p . So D_p contains a left ideal and a right ideal each with at least $|A|$ elements. They cannot have more than $|A|$ elements because for each $F \in \mathcal{P}_f(\lambda)$ with $p = \max F$, there are $|A|^{|F|} = |A|$ choices for $\prod_{q \in F} g(x_q)$. So $|D_p| = |A|$. Thus $D_p \approx L \times R \approx V_{|A|}$. \square

Two obvious questions are raised by Lemma 3.4 and Theorem 3.5. First, can the function g constructed there be required to be continuous? Second, can the function g in Theorem 3.5 be required to be a homomorphism? We shall answer both of these questions in the negative, even when the stronger requirements that T and C be compact and C be a *topological* semigroup are added. We shall have need of the following lemma, whose routine proof we omit. (Recall that any successor ordinal is a compact Hausdorff space under its order topology.)

Lemma 3.6. *Let A be a compact space, let λ be an ordinal, let $A \times A \times (\lambda + 1)$ have the product topology, and let A and $A \times A \times (\lambda + 1)$ be clopen subsets of $C = C_{A,\lambda}$. Then C is a compact topological semigroup and $C_\lambda = K(C)$.*

We now show that, even for $\lambda = 0$, one cannot require that g be continuous. We remind the reader that an F -space is a completely regular space X in which $\{x \in X : f(x) > 0\}$ and $\{x \in X : f(x) < 0\}$ are completely separated for all continuous $f : X \rightarrow \mathbb{R}$.

Theorem 3.7. *There exist a nonempty set A , a topology on $C = C_{A,1}$ such that C is a compact topological semigroup and A and C_0 are compact subsets of C , a compact right topological semigroup T , and a continuous surjective homomorphism $f : T \rightarrow C$ such that there is no continuous homomorphism $g : C \rightarrow T$ for which $f \circ g$ is the identity on C . (In fact, there is no continuous injective function from C to T .)*

Proof. Let $A = \beta\mathbb{N}$, let $C = \beta\mathbb{N} \cup (\beta\mathbb{N} \times \beta\mathbb{N} \times \{0\})$ with the topology given in Lemma 3.6, and let $T = \mathbb{H}_\omega$. Then $\mathbb{N} \cup (\mathbb{N} \times \mathbb{N} \times \{0\})$ is dense in $C = \Lambda(C)$. So there is a continuous surjective homomorphism $f : T \rightarrow C$ by Theorem 2.5.

Now suppose there is a continuous injective function $g : C \rightarrow T$. Then by Theorem 2.2 there is a continuous injective function from C to $\mathbb{H} \subseteq \beta\mathbb{N}$. But this is impossible because $\beta\mathbb{N}$ is an F -space [3, Theorem 14.25]. So every compact subset X of $\beta\mathbb{N}$ is an F -space, because every continuous function from X to $[0,1]$ has a continuous extension to $\beta\mathbb{N}$, by the Tietze extension theorem. But $\beta\mathbb{N} \times \beta\mathbb{N}$ is not an F -space by [3, 14Q]. \square

Theorem 3.8. *There exist a nonempty set A with distinguished element 0, a topology on $C = C_{A,\omega+1}$ such that C is a compact topological semigroup and A and C_p are compact subsets of C for each $p \leq \omega$, a compact right topological semigroup T , and a continuous surjective homomorphism $f : T \rightarrow C$ such that there is no homomorphism $g : C \rightarrow T$ for which $f \circ g$ is the identity on C .*

Proof. Let $A = \{0\}$ and let $C = C_{A,\omega+1}$ with the topology given in Lemma 3.6. Let $u_0 = 0$, for $p < \omega$, let $u_{p+1} = (0, 0, p)$, and let $u_\omega = (0, 0, \omega)$. Then $C = \{u_p : p \leq \omega\}$ and $u_p u_q = u_{p \vee q}$ for all $p, q \leq \omega$. Topologically, u_ω is the only non-isolated point in C . Let $\langle v_p \rangle_{p < \omega}$ be a sequence of distinct points none of which are in C . Let $T = \{u_p : p < \omega\} \cup \{v_p : p < \omega\}$ and for $p, q < \omega$ define an operation on T as follows:

$$\begin{aligned} u_p u_q &= u_{p \vee q}, \\ u_p v_q &= v_{p \vee q}, \\ v_p u_q &= v_p v_q = v_p. \end{aligned}$$

We leave it to the reader to verify that the operation is associative.

Let $T \setminus \{v_0\}$ be discrete, and let T be the one point compactification of $T \setminus \{v_0\}$ (with v_0 as the point at infinity). We claim that T is a right topological semigroup. Let $p < \omega$. To see that ρ_{u_p} is continuous at v_0 , let W be a neighborhood of $v_0 = \rho_{u_p}(v_0)$ and let $U = W \cap (\{u_q : p \leq q < \omega\} \cup \{v_q : q < \omega\})$. Then $\rho_{u_p}[U] \subseteq W$. To see that ρ_{v_p} is continuous at v_0 , let W be a neighborhood of $v_0 = \rho_{v_p}(v_0)$ and let $U = \{u_q : p \leq q < \omega \text{ and } v_q \in W\} \cup \{v_q : v_q \in W\}$. Then $\rho_{v_p}[U] \subseteq W$.

Define $f : T \rightarrow C$ by $f(u_p) = u_p$ and $f(v_p) = u_\omega$ for each $p < \omega$. Then f is a continuous surjective homomorphism. Suppose that $g : C \rightarrow T$ is a homomorphism for which $f \circ g$ is the identity. Then for $p < \omega$, $g(u_p) = u_p$, and there is some $q < \omega$ such that $g(u_\omega) = v_q$. But then, $v_{q+1} = u_{q+1} v_q = g(u_{q+1}) \cdot g(u_\omega) = g(u_{q+1} u_\omega) = g(u_\omega) = v_q$, a contradiction. \square

We shall see next that we can get the function g to be a homomorphism by requiring that T be semitopological. This corollary can then be viewed as saying that C is something like an absolute co-retract in the category of semitopological semigroups. C becomes an absolute co-retract in the category of compact semitopological semigroups if it is given a topology for which it is in this category with A and each C_p being compact.

Corollary 3.9. *Let A be a nonempty set with distinguished element 0, let λ be an ordinal, and let $C = C_{A,\lambda}$. Let T be a semitopological semigroup, and $f : T \rightarrow C$ be a continuous homomorphism such that $f^{-1}[A]$ is compact and $f^{-1}[C_p]$ is compact for every $p < \lambda$. Then there is a homomorphism $g : C \rightarrow T$ such that $f \circ g$ is the identity.*

Proof. At stage p of the induction in the proof of Theorem 3.5 one has that for each $q < p$, $g(u_q) \cdot f^{-1}[c_p]$ is a compact right ideal of $f^{-1}[c_p]$. So one may choose a minimal right ideal R of $f^{-1}[c_p]$ with $R \subseteq \bigcap_{q < p} g(u_q) \cdot f^{-1}[c_p]$ and $f[R] = u_p \cdot C_p$. Then, if $y \in C_q$ for some $q \leq p$ and $x \in C_p$, just as one showed in the verification of hypothesis (i) that $g(x) \cdot g(y) \in L \cdot g(y_r)$, one can show that $g(y) \cdot g(x) \in g(y_l) \cdot R$, so that $g(y) \cdot g(x)$ and $g(yx)$ are idempotents in the same group. If $y \in A$ and $x \in C_p$, then $g(y) \cdot g(x) = g(yx)$ by (ii). (We did not need to consider the case $y \in A$ separately at that point in the proof of Theorem 3.5 because the equation $u_q x_r y = u_q y_r$ was valid in any event. The corresponding equation $yx_l u_q = y_l u_q$ is not valid if $y \in A$.) \square

We now present some immediate consequences of Theorem 3.5, although with a bit more effort, we shall get a stronger result, namely Theorem 3.16.

Corollary 3.10. *Let κ be an infinite cardinal and let λ be an ordinal with $|\lambda| \leq \kappa$. Then \mathbb{H}_κ contains a subsemigroup $D = \bigcup_{p \leq \lambda} D_p$ of idempotents where each D_p is a rectangular component of D isomorphic to $V_{2^{2^\kappa}}$ and the sequence $\langle D_p \rangle_{p \leq \lambda}$ is decreasing in the ordering of components, so that for each $p \leq \lambda$, $D_p = K(\bigcup_{q \leq p} D_q)$.*

Proof. Let κ have the discrete topology and let $A = \beta\kappa$. Let $C = C_{A, \lambda+1}$ and let C have the topology described in Lemma 3.6. Let $T = \mathbb{H}_\kappa$. Since $\kappa \cup (\kappa \times \kappa \times (\lambda+1))$ is a dense subset of $C = \Lambda(C)$, by Theorem 2.5 there is a continuous surjective homomorphism $f : T \rightarrow C$, and so Theorem 3.5 applies. \square

Corollary 3.11. *Let S be an infinite cancellative discrete semigroup with cardinality κ and let λ be an ordinal with $|\lambda| \leq \kappa$. Then $\beta S \setminus S$ contains a subsemigroup $D = \bigcup_{p \leq \lambda} D_p$ of idempotents where each D_p is a rectangular component of D isomorphic to $V_{2^{2^\kappa}}$ and the sequence $\langle D_p \rangle_{p \leq \lambda}$ is decreasing in the ordering of components, so that for each $p \leq \lambda$, $D_p = K(\bigcup_{q \leq p} D_q)$. If $S = (\mathbb{N}, +)$, $S = (\mathbb{N}, \cdot)$, or S is a countably infinite discrete group, then $D_\lambda \subseteq K(\beta S)$.*

Proof. By Theorem 2.7, $\beta S \setminus S$ contains a topological and algebraic copy T of \mathbb{H}_κ . (If $S = (\mathbb{N}, +)$, choose $T = \mathbb{H}$. If $S = (\mathbb{N}, \cdot)$ or S is a countably infinite discrete group, choose T as in Theorem 2.4 or Theorem 2.3 respectively.) Then Corollary 3.10 applies.

If $S = (\mathbb{N}, +)$, $S = (\mathbb{N}, \cdot)$, or S is a countably infinite discrete group, then $K(T) = K(\beta S) \cap T$. So by Theorem 3.5, with $\lambda + 1$ in place of λ , we have $g[C_\lambda] \subseteq K(T) \subseteq K(\beta S)$ and so $D_\lambda \subseteq K(\beta S)$. \square

Corollary 3.12. *Let S be a countably infinite discrete group. Then there is a copy of V_{2^c} contained in $K(\beta S)$.*

We can completely characterise the semigroups of idempotents that can be embedded in $K(\beta\mathbb{N})$.

Corollary 3.13. *Let S be an infinite cancellative discrete semigroup with cardinality κ and let D be a semigroup of idempotents.*

- (i) *There is a copy of D in $\beta S \setminus S$ if D is rectangular and $|D| \leq 2^{2^\kappa}$.*
- (ii) *There is a copy of D in $K(\beta\mathbb{N})$ if and only if D is rectangular and $|D| \leq 2^c$.*

Proof. Conclusion (i) and the sufficiency of (ii) follow immediately from Corollary 3.11. Assume now that D is a semigroup of idempotents contained in $K(\beta\mathbb{N})$. Then $|D| \leq |\beta\mathbb{N}| = 2^c$. Next observe that any subsemigroup of idempotents in $K(\beta S)$

must be rectangular. To see this, suppose that $x, y, z \in K(\beta S)$. Then xz and xyz belong to the same minimal left ideal and to the same minimal right ideal. Hence, if they are idempotent, they must be equal. \square

Recall that any two maximal groups in the smallest ideal of a compact right topological semigroup are isomorphic. We see that we can get the direct product of such groups with an embedded rectangular semigroup in the smallest ideal as well.

Theorem 3.14. *Let T be a compact right topological semigroup, let D be a rectangular subsemigroup of $K(T)$, and let G be a maximal subgroup of $K(T)$. There is an algebraic copy of $D \times G$ contained in $K(T)$.*

Proof. Let L be a minimal left ideal of D and let R be a minimal right ideal of D . Since D is rectangular, D is the internal direct product of L and R , meaning that each element x of D can be written uniquely as $x = x_L x_R$ where $x_L \in L$ and $x_R \in R$. Also, $RL = R \cap L$ is a subgroup of D and so, since D consists of idempotents, $RL = \{e\}$ for some e . Then for any $x, y \in D$, $x_R y_L = e$. Note also that $(xy)_L = x_L$ and $(xy)_R = y_R$.

We may assume that $G = eTe$. Define $\varphi : D \times G \rightarrow K(T)$ by $\varphi(x, g) = x_L g x_R$. We claim that φ is an injective homomorphism. Let $(x, g), (y, h) \in D \times G$. Then $\varphi((x, g)(y, h)) = x_L g x_R y_L h y_R = x_L g e h y_R = (xy)_L g h (xy)_R = \varphi(xy, gh)$.

Now assume that $\varphi(x, g) = \varphi(y, h)$. Then $g = ege = x_R x_L g x_R x_L = x_R y_L h y_R x_L = e h e = h$. Also, $x_L T \cap y_L T \neq \emptyset$ and $x_L T$ and $y_L T$ are minimal right ideals of T ; so $x_L T = y_L T$. Similarly $T x_R = T y_R$. Now $x = x_L x_R \in x_L T \cap T x_R$ and $y \in y_L T \cap T y_R$. So x and y are idempotents in the same group and therefore $x = y$. \square

Corollary 3.15. *$K(\beta \mathbb{N})$ contains an algebraic copy of $V_{2^c} \times F$, where V_{2^c} is the $2^c \times 2^c$ rectangular semigroup and F is the free group on 2^c generators.*

Proof. By Corollary 3.13, $K(\beta \mathbb{N})$ contains a copy of the $2^c \times 2^c$ rectangular semigroup, and by [4], each maximal group in $K(\beta \mathbb{N})$ contains a copy of the free group on 2^c generators. Therefore the result follows from Theorem 3.14. \square

We now present a strengthening of Corollary 3.10, producing a longer chain of rectangular components. Recall that the *Souslin number* $S(X)$ of a topological space X (also known as the *cellularity* of X) is the least cardinal γ such that X does not have a collection of γ pairwise disjoint nonempty open subsets. See [2, Chapter 12] for considerable information about the Souslin number of the space $U(\kappa)$ of uniform ultrafilters on κ . Recall, in particular, that the Souslin number of $\mathbb{N}^* = \beta \mathbb{N} \setminus \mathbb{N} = U(\mathbb{N})$ is \mathfrak{c}^+ .

Theorem 3.16. *Let κ be an infinite cardinal and let λ be an infinite ordinal for which $|\lambda| < S(U(\kappa))$. There exist a set A with $|A| = 2^{2^\kappa}$ and an injection $g : C_{A, \lambda} \rightarrow \mathbb{H}_\kappa$ such that if $q < p < \lambda$, $y \in C_q$, and $x \in C_p$, then $g(x) \cdot g(y) = g(xy)$, and if $p = q + n$ for some $n < \omega$, then $g(y) \cdot g(x) = g(yx)$. Also, \mathbb{H}_κ contains a semigroup $D = \bigcup_{p < \lambda} D_p$ of idempotents where for each $p < \lambda$, D_p is a rectangular component of D isomorphic to $V_{2^{2^\kappa}}$, $g[C_p] \subseteq D_p$, and the sequence $\langle D_p \rangle_{p < \lambda}$ is decreasing in the ordering of components, so that for each $p < \lambda$, $D_p = K(\bigcup_{q \leq p} D_q)$. If λ is a successor, then $D_{\lambda-1} \subseteq K(\mathbb{H}_\kappa)$.*

Proof. Since $|\lambda| < S(U(\kappa))$, choose a family $\langle E_p \rangle_{p < \lambda}$ of subsets of κ such that each $|E_p| = \kappa$ and $|E_p \cap E_q| < \kappa$ when $p \neq q$. For each $p < \lambda$ we define $\phi_p : W_\kappa \rightarrow W_\kappa$

by $\phi_p(w) = \sum_{\alpha \in E_p \cap \text{supp}(w)} e_\alpha$ (where $\sum_{\alpha \in \emptyset} e_\alpha = 0$) and let $\widetilde{\phi}_p : \beta W_\kappa \rightarrow \beta W_\kappa$ be the continuous extension of ϕ_p . If $v, w \in W_\kappa$ and $\text{supp}(v) \cap \text{supp}(w) = \emptyset$, then $\phi_p(v + w) = \phi_p(v) + \phi_p(w)$. So by [6, Theorem 4.21] the restriction of $\widetilde{\phi}_p$ to \mathbb{H}_κ is a homomorphism.

Next observe that for $x \in \mathbb{H}_\kappa$, $\widetilde{\phi}_p(x) \in \{0\} \cup \mathbb{H}_\kappa$. If there exist $B \in x$ and $\alpha < \kappa$ such that $\text{supp}(w) \cap E_p = \emptyset$ whenever $w \in B$ and $\min \text{supp}(w) \geq \alpha$, then $\widetilde{\phi}_p(x) = 0$ because ϕ_p is constantly 0 on $\{w \in B : \min \text{supp}(w) \geq \alpha\} \in x$. Otherwise $\{\phi_p[\{w \in B : \min \text{supp}(w) \geq \alpha\}] : B \in x \text{ and } \alpha < \kappa\}$ has the finite intersection property and so is contained in an ultrafilter y . This $y \in \mathbb{H}_\kappa$ and $y = \widetilde{\phi}_p(x)$.

Let $T_0 = \mathbb{H}_\kappa \cap \text{cl}\{w \in W_\kappa : \text{supp}(w) \subseteq E_0\}$. Notice that T_0 is a compact subsemigroup of \mathbb{H}_κ . For each p with $0 < p < \lambda$ let

$$T_p = \{x \in \mathbb{H}_\kappa : \widetilde{\phi}_p(x) \in \mathbb{H}_\kappa \text{ and for all } q \text{ with } p < q \leq \lambda, \widetilde{\phi}_q(x) = 0\}.$$

To see that $T_p \neq \emptyset$, let x be a uniform ultrafilter on $\{e_\alpha : \alpha \in E_p\}$. If $q \neq p$, then $|E_q \cap E_p| < \kappa$. So $\widetilde{\phi}_q(x) = 0$, while $\widetilde{\phi}_p(x) = x \in \mathbb{H}_\kappa$ (because ϕ_p is the identity on $\{e_\alpha : \alpha \in E_p\}$). Since $\widetilde{\phi}_q$ is a homomorphism on \mathbb{H}_κ for each $q \leq \lambda$ we have that T_p is a subsemigroup of \mathbb{H}_κ . Since $T_p = \mathbb{H}_\kappa \cap \widetilde{\phi}_p^{-1}[\mathbb{H}_\kappa] \cap \bigcap_{p < q \leq \lambda} \widetilde{\phi}_q^{-1}[\{0\}]$, T_p is compact.

If λ is a successor, let $T_{\lambda-1} = \mathbb{H}_\kappa \cap \bigcap_{p < \lambda} \widetilde{\phi}_p^{-1}[\mathbb{H}_\kappa]$. Then $T_{\lambda-1}$ is clearly a compact subsemigroup of \mathbb{H}_κ provided $T_{\lambda-1} \neq \emptyset$. We show in fact that $K(\mathbb{H}_\kappa) \subseteq T_{\lambda-1}$. Let $x \in K(\mathbb{H}_\kappa)$, let $p < \lambda$, and let y be a uniform ultrafilter on $\{e_\alpha : \alpha \in E_p\}$. By [6, Theorem 4.39] pick $z \in \mathbb{H}_\kappa$ such that $x = z + y + x$. Then $\widetilde{\phi}_p(x) = \widetilde{\phi}_p(z) + \widetilde{\phi}_p(y) + \widetilde{\phi}_p(x) = \widetilde{\phi}_p(z) + y + \widetilde{\phi}_p(x) \neq 0$.

Next observe that for $p, q < \lambda$, $T_p + T_q = T_{p \vee q}$ and if $p \neq q$, then $T_p \cap T_q = \emptyset$.

Let $T = \bigcup_{p < \lambda} T_p$. If T has the relative topology induced by \mathbb{H} , T is a right topological semigroup.

Let $A = \mathbb{H}_\kappa \cap \text{cl}\{e_\alpha : \alpha \in E_0\}$. Then A is exactly the set of uniform ultrafilters on $\{e_\alpha : \alpha \in E_0\}$, and so $|A| = 2^{2^\kappa}$. Let $C = C_{A, \lambda}$.

We shall now construct a surjective homomorphism $f : T \rightarrow C$. We first introduce some mappings. Let $\theta : W_\kappa \rightarrow \{e_\alpha : \alpha \in E_0\}$ be a function whose restriction to $\{e_\alpha : \alpha \in E_0\}$ is the identity, whose restriction to $\{e_\alpha : \alpha \in E_1\}$ is a bijection, and whose restriction to $W_\kappa \setminus \{e_\alpha : \alpha \in E_0 \cup E_1\}$ is a bijection. (In particular, θ is at most three-to-one.) Let $\widetilde{\theta} : \beta W_\kappa \rightarrow \text{cl}\{e_\alpha : \alpha \in E_0\}$ be the continuous extension of θ .

Let $\epsilon(0) = \delta(0) = 0$. For $w \in W_\kappa \setminus \{0\}$, let $\epsilon(w) = e_\gamma$ where $\gamma = \max \text{supp}(w)$. If $\text{supp}(w) \subseteq E_0$, let $\delta(w) = 0$. Otherwise let $\delta(w) = e_\alpha$ where $\alpha = \min(\text{supp}(w) \setminus E_0)$. Let $\widetilde{\delta} : \beta W_\kappa \rightarrow \{0\} \cup \text{cl}\{e_\alpha : \alpha < \kappa\}$ and $\widetilde{\epsilon} : \beta W_\kappa \rightarrow \{0\} \cup \text{cl}\{e_\alpha : \alpha < \kappa\}$ be the continuous extensions of δ and ϵ respectively. Notice that $\widetilde{\delta}$ is the identity on $\{e_\alpha : \alpha \in \kappa \setminus E_0\}$ and $\widetilde{\epsilon}$ is the identity on $\{e_\alpha : \alpha < \kappa\}$. So $\widetilde{\delta}$ is the identity on $\mathbb{H}_\kappa \cap \text{cl}\{e_\alpha : \alpha \in \kappa \setminus E_0\}$ and $\widetilde{\epsilon}$ is the identity on $\mathbb{H}_\kappa \cap \text{cl}\{e_\alpha : \alpha < \kappa\}$. We claim that for $x, y \in \mathbb{H}_\kappa$,

$$\widetilde{\epsilon}(x + y) = \widetilde{\epsilon}(y),$$

$$(*) \quad \widetilde{\delta}(x + y) = \begin{cases} \widetilde{\delta}(x) & \text{if } x \notin T_0, \\ \widetilde{\delta}(y) & \text{if } x \in T_0. \end{cases}$$

For $w, v \in W_\kappa$, if $\max \text{supp}(v) > \max \text{supp}(w)$, then $\epsilon(w + v) = \epsilon(v)$ so that $\tilde{\epsilon}(w + y) = \tilde{\epsilon}(y)$; if $\text{supp}(w) \subseteq E_0$, then $\delta(w + v) = \delta(v)$ so that $\tilde{\delta}(w + y) = \tilde{\delta}(y)$. For $w, v \in W_\kappa$, if $\max \text{supp}(w) < \min \text{supp}(v)$ and $\text{supp}(w) \setminus E_0 \neq \emptyset$, then $\delta(w + v) = \delta(w)$; so $\tilde{\delta}(w + y) = \tilde{\delta}(w)$. The equations in (*) then follow by the continuity of $\tilde{\delta} \circ \rho_y$ and $\tilde{\epsilon} \circ \rho_y$.

For $x \in T_0$, let $f(x) = \tilde{\theta}(\tilde{\epsilon}(x))$. If $0 < p < \lambda$ and $x \in T_p$, let

$$f(x) = \begin{cases} (\tilde{\theta}(\tilde{\delta}(x)), \tilde{\theta}(\tilde{\epsilon}(x)), p-1) & \text{if } p < \omega, \\ (\tilde{\theta}(\tilde{\delta}(x)), \tilde{\theta}(\tilde{\epsilon}(x)), p) & \text{if } p \geq \omega. \end{cases}$$

For any $x \in \mathbb{H}_\kappa$, one has $\tilde{\theta}(\tilde{\delta}(x)) \in A$ and $\tilde{\theta}(\tilde{\epsilon}(x)) \in A$. (We have that $\tilde{\theta}[\mathbb{H}_\kappa] \subseteq \mathbb{H}_\kappa$ because θ is at most three-to-one.) Thus $f[T_0] \subseteq A$, $f[T_p] \subseteq C_{p-1}$ if $0 < p < \omega$, and $f[T_p] \subseteq C_p$ if $\omega \leq p < \lambda$.

Given $x \in A \subseteq T_0$, one has $f(x) = \tilde{\theta}(\tilde{\epsilon}(x)) = \tilde{\theta}(x) = x$; so $f[T_0] = A$. Now let $p < \lambda$ and let $(y, z, p) \in C_p$. If $p < \omega$, let $q = p + 1$; otherwise let $q = p$. Pick $y' \in \text{cl}\{e_\alpha : \alpha \in E_1\}$ such that $\tilde{\theta}(y') = y$. Pick $x \in T_q$. Then $y' + x + z \in T_q$ and

$$\begin{aligned} f(y' + x + z) &= (\tilde{\theta}(\tilde{\delta}(y' + x + z)), \tilde{\theta}(\tilde{\epsilon}(y' + x + z)), p) \\ &= (\tilde{\theta}(\tilde{\delta}(y')), \tilde{\theta}(\tilde{\epsilon}(z)), p) \\ &= (\tilde{\theta}(y'), \tilde{\theta}(z), p) \\ &= (y, z, p). \end{aligned}$$

Therefore, $f[T_q] = C_p$.

The verification that f is a homomorphism is routine using the equations (*).

Choose $g : C \rightarrow T$ and $\langle D_p \rangle_{p < \lambda}$ as guaranteed by Theorem 3.5. Since we have already observed that $|A| = 2^{2^\kappa}$, all conclusions follow immediately except the assertion that $D_{\lambda-1} \subseteq K(\mathbb{H}_\kappa)$ when λ is a successor. To see this recall that $K(\mathbb{H}_\kappa) \subseteq T_{\lambda-1}$ so that $K(\mathbb{H}_\kappa)$ is an ideal of $T_{\lambda-1}$ and thus $K(T_{\lambda-1}) \subseteq K(\mathbb{H}_\kappa)$. By Theorem 3.5(iv), $g[C_{\lambda-1}] \subseteq K(f^{-1}[C_{\lambda-1}]) = K(T_{\lambda-1})$. So $D \cap K(\mathbb{H}_\kappa) \neq \emptyset$ and is thus an ideal of D and therefore $D_{\lambda-1} = K(D) \subseteq D \cap K(\mathbb{H}_\kappa)$. \square

Corollary 3.17. *Let λ be an ordinal for which $|\lambda| = \mathfrak{c}$. There exist a set A with $|A| = 2^{\mathfrak{c}}$ and an injection $g : C_{A,\lambda} \rightarrow \mathbb{H}$ such that if $q < p < \lambda$, $y \in C_q$, and $x \in C_p$, then $g(x) \cdot g(y) = g(xy)$, and if $p = q + n$ for some $n < \omega$, then $g(y) \cdot g(x) = g(yx)$. Also, \mathbb{H} contains a semigroup $D = \bigcup_{p < \lambda} D_p$ of idempotents where for each $p < \lambda$, D_p is a rectangular component of D , $g[C_p] \subseteq D_p$, and the sequence $\langle D_p \rangle_{p < \lambda}$ is decreasing in the ordering of components. For each $p < \lambda$, $|D_p| = 2^{\mathfrak{c}}$ and if λ is a successor, then $D_{\lambda-1} \subseteq K(\mathbb{H}) \subseteq K(\beta\mathbb{N})$.*

Proof. By Theorem 2.2, \mathbb{H} and \mathbb{H}_ω are topologically and algebraically isomorphic. Also $S(U(\omega)) = \mathfrak{c}^+$. So this is an immediate consequence of Theorem 3.16. \square

It was shown in [5, Corollary 3.4] that there is a \leq_L -chain $\langle u_\sigma \rangle_{\sigma < \omega_1}$ of distinct idempotents in $\beta\mathbb{N}$ with the property that for each $\sigma < \omega_1$, $u_{\sigma+1} \leq u_\sigma$. We are now able to establish a considerably stronger statement. (The necessity in the following corollary was also established in [5], but we include the short proof for completeness.)

Corollary 3.18. *Let λ be an ordinal. There is a \leq_L -chain $\langle u_\sigma \rangle_{\sigma < \lambda}$ of distinct idempotents in $\beta\mathbb{N}$ with the property that for each $\sigma < \lambda$, $u_{\sigma+1} \leq u_\sigma$ if and only*

if $|\lambda| \leq \mathfrak{c}$. If $|\lambda| \leq \mathfrak{c}$ and λ is a successor, one may choose such a sequence with $u_{\lambda-1} \in K(\beta\mathbb{N})$.

Proof. Necessity. For each $\sigma < \lambda$, $\mathbb{N}^* + u_\sigma$ properly contains the compact set $\mathbb{N}^* + u_{\sigma+1}$. So one can choose a clopen subset U_σ of $\beta\mathbb{N}$ with $\mathbb{N}^* + u_{\sigma+1} \subseteq U_\sigma$ and $(\mathbb{N}^* + u_\sigma) \setminus U_\sigma \neq \emptyset$. The clopen subsets of $\beta\mathbb{N}$ correspond exactly to the subsets of \mathbb{N} and so there are exactly \mathfrak{c} of them.

Sufficiency. Choose A and g as guaranteed by Corollary 3.17 for λ . For each $p < \lambda$, let $u_p = g(0, 0, p)$. If λ is a successor, then $u_{\lambda-1} \in g[C_{\lambda-1}] \subseteq D_{\lambda-1} \subseteq K(\beta\mathbb{N})$. \square

Question 3.19. Is there a decreasing \leq -chain of idempotents in \mathbb{N}^* indexed by $\omega + 1$?

We close this section by observing that it is consistent with ZFC that there are idempotents in $\beta\mathbb{N}$ that are not members of any nontrivial rectangular subsemigroup of $\beta\mathbb{N}$. Indeed, by [6, Theorems 12.19 and 12.29 and Lemma 12.44], Martin’s Axiom implies that there is an idempotent $p \in \beta\mathbb{N}$ such that, whenever $q \in \beta\mathbb{N}$, $r \in \bigcap_{n=1}^\infty \text{cl}_{\beta\mathbb{N}}(\mathbb{N}n)$, and $p = q + r$, one must have $p = q = r$. In particular, if $p = p + q + p$, then $p = q$.

It can be shown in ZFC that there are idempotents p in $\beta\mathbb{N}$ that are strongly right maximal; i.e., the equation $q + p = p$, with $q \in \beta\mathbb{N}$, implies that $q = p$ [6, Theorem 9.10]. If p is an idempotent of this kind, p does not belong to any semigroup in $\beta\mathbb{N}$ isomorphic to a semigroup of the form $V_{|A|}$ unless $|A| = 1$.

4. CHAINS OF RECTANGULAR SEMIGROUPS AS CO-RETRACTS

It was shown in [10] that certain infinite chains of finite rectangular semigroups are absolute co-retracts. We prove in this section a similar theorem in which the rectangular semigroups are allowed to be infinite. As a consequence, we obtain additional semigroups which can be algebraically embedded in \mathbb{H}_κ .

Definition 4.1. Let $\mathcal{A} = \langle A_n \rangle_{n < \omega}$ and $\mathcal{B} = \langle B_n \rangle_{n < \omega}$ be sequences of sets. Assume that each A_n has a designated element α_n and each B_n has a designated element δ_n . Suppose also that, for each $n < \omega$, either $A_n = \{\alpha_n\}$ or $B_n = \{\delta_n\}$. For each $p < \omega$, we define D_p to be the set of pairs of words of the form $(a_0 a_1 \cdots a_p, b_p b_{p-1} \cdots b_0)$, where $a_i \in A_i$ and $b_i \in B_i$ for each $i \in \{0, 1, \dots, p\}$. For $0 < \lambda \leq \omega$, we let $D_{\mathcal{A}, \mathcal{B}, \lambda} = \bigcup_{p < \lambda} D_p$. We define a semigroup operation on $D_{\mathcal{A}, \mathcal{B}, \lambda}$ as follows: if $x = (a_0 a_1 \cdots a_p, b_p b_{p-1} \cdots b_0) \in D_p$ and $y = (c_0 c_1 \cdots c_q, d_q d_{q-1} \cdots d_0) \in D_q$, where $q \leq p$, then

$$xy = (a_0 a_1 \cdots a_p, b_p \cdots b_{q+1} d_q d_{q-1} \cdots d_0)$$

and

$$yx = (c_0 c_1 \cdots c_q a_{q+1} \cdots a_p, b_p b_{p-1} \cdots b_0).$$

We leave the verification that the operation is associative to the reader. Observe that each D_p is a rectangular semigroup.

Notice that if A is a set with designated element 0, $A_0 = \{0\}$, $B_0 = A$, $A_1 = A$, $B_1 = \{0\}$, and $A_n = B_n = \{0\}$ for $n > 1$, then $D_{\mathcal{A}, \mathcal{B}, \omega}$ is isomorphic to $C_{A, \omega}$. (Send $(0, a)$ to a and for $p > 0$ send the element $(0a00 \cdots 0, 00 \cdots 0b)$ of D_p to $(a, b, p-1)$.) Thus the structure of $D_{\mathcal{A}, \mathcal{B}, \omega}$ is, in general, considerably more complicated than that of $C_{A, \omega}$.

Definition 4.2. Let $p < \omega$ and let $x = (a_0 a_1 \cdots a_p, b_p b_{p-1} \cdots b_0) \in D_p$. We define elements $\phi_1(x)$ and $\phi_2(x)$ in D_p by $\phi_1(x) = (\alpha_0 \alpha_1 \cdots \alpha_{p-1} a_p, \delta_p \delta_{p-1} \cdots \delta_0)$ and $\phi_2(x) = (\alpha_0 \alpha_1 \cdots \alpha_p, b_p \delta_{p-1} \delta_{p-2} \cdots \delta_0)$ and if $p > 0$, we define x_ℓ and x_r in D_{p-1} by $x_\ell = (a_0 a_1 \cdots a_{p-1}, \delta_{p-1} \delta_{p-2} \cdots \delta_0)$ and $x_r = (\alpha_0 \alpha_1 \cdots \alpha_{p-1}, b_{p-1} b_{p-2} \cdots b_0)$. We put $u_p = (\alpha_0 \alpha_1 \cdots \alpha_p, \delta_p \delta_{p-1} \cdots \delta_0) \in D_p$.

We show that D is something like an absolute co-retract.

Theorem 4.3. Let $\mathcal{A} = \langle A_n \rangle_{n < \omega}$ and $\mathcal{B} = \langle B_n \rangle_{n < \omega}$ be sequences of sets as in Definition 4.1, let $0 < \lambda \leq \omega$, and let $D = D_{\mathcal{A}, \mathcal{B}, \lambda}$. Let T be a right topological semigroup, and let $f : T \rightarrow D$ be a surjective homomorphism such that $f^{-1}[D_p]$ is compact for each $p < \lambda$. Then there is a homomorphism $g : D \rightarrow T$ for which $f \circ g$ is the identity. If T is compact and $\lambda < \omega$, then $g[D_{\lambda-1}] \subseteq K(T)$.

Proof. We may assume that $A_0 = \{\alpha_0\}$ so that D_0 is a right zero semigroup. Exactly as in the first paragraph of the proof of Lemma 3.4 we can define $g : D_0 \rightarrow T$ such that g is a homomorphism and $f \circ g$ is the identity on D_0 . So we assume that $p > 0$ and g has been defined on $\bigcup_{q < p} D_q$.

For each $x \in D_p$, note that $x D_p$ is a minimal right ideal of D_p and $D_p x$ is a minimal left ideal of D_p . So we may choose a minimal right ideal $R(x)$ of $f^{-1}[D_p]$ and a minimal left ideal $L(x)$ of $f^{-1}[D_p]$ such that $f[R(x)] = x D_p$ and $f[L(x)] = D_p x$. Given $x \in D_p$, we have by Lemma 3.3 that $g(x_\ell) R(\phi_1(x))$ is a minimal right ideal of $f^{-1}[D_p]$ and $L(\phi_2(x)) g(x_r)$ is a minimal left ideal of $f^{-1}[D_p]$. So we may define $g(x)$ to be the identity of the group $g(x_\ell) R(\phi_1(x)) L(\phi_2(x)) g(x_r)$. Notice that if T has a smallest ideal (in particular if T is compact) and $\lambda = p + 1$, then $K(T) \subseteq f^{-1}[D_p]$. So $K(f^{-1}[D_p]) \subseteq K(T)$ and thus $g[D_p] \subseteq K(T)$.

Now $f(g(x)) \in x_\ell \phi_1(x) D_p = x D_p$ and $f(g(x)) \in D_p \phi_2(x) x_r = D_p x$. So $f(g(x))$ is an idempotent in $x D_p x$ and thus $f(g(x)) = x$.

Suppose that $x \in D_p$ and $y \in D_q$ where $q \leq p$. Then $\phi_1(x) = \phi_1(xy)$ and $x_\ell = (xy)_\ell$. So $g(xy) \in g(x_\ell) R(\phi_1(x))$ and $g(x) g(y) \in g(x_\ell) R(\phi_1(x)) g(y) \subseteq g(x_\ell) R(\phi_1(x))$ by Lemma 3.3. Therefore, $g(xy)$ and $g(x) g(y)$ are members of the same minimal right ideal of $f^{-1}[D_p]$.

If $q < p$, $\phi_2(xy) = \phi_2(x)$ and $(xy)_r = x_r y$. So $g(xy) \in L(\phi_2(x)) g((xy)_r)$ and $g(x) g(y) \in L(\phi_2(x)) g(x_r) g(y) = L(\phi_2(x)) g(x_r y) = L(\phi_2(x)) g((xy)_r)$.

If $q = p$, $\phi_2(xy) = \phi_2(y)$ and $(xy)_r = y_r$. So $g(xy) \in L(\phi_2(y)) g(y_r)$ and $g(x) g(y) \in g(x) L(\phi_2(y)) g(y_r)$. Thus in any event $g(xy)$ and $g(x) g(y)$ are in the same minimal left ideal of $f^{-1}[D_p]$.

By a left-right switch of the above arguments we have that $g(yx)$ and $g(y) g(x)$ are in the same minimal left ideal and the same minimal right ideal of $f^{-1}[D_p]$.

First assume that $q < p$. Pick $a \in R(\phi_1(x)) L(\phi_2(x))$ such that $g(x) = g(x_\ell) a g(x_r)$. Then

$$\begin{aligned} g(x) g(y) g(x) &= g(x_\ell) a g(x_r) g(y) g(x_\ell) a g(x_r) \\ &= g(x_\ell) a g(x_r y x_\ell) a g(x_r) \\ &= g(x_\ell) a g(x_r x_\ell) a g(x_r) \\ &= g(x_\ell) a g(x_r) g(x_\ell) a g(x_r) = g(x) g(x) = g(x). \end{aligned}$$

So $g(x) g(y) g(x) g(y) = g(x) g(y)$ and $g(y) g(x) g(y) g(x) = g(y) g(x)$ and thus $g(x) g(y) = g(xy)$ and $g(y) g(x) = g(yx)$.

Now assume that $q = p$. Assume also that $B_p = \{\delta_p\}$. (The case that $A_p = \{\alpha_p\}$ then proceeds by a left-right switch of the following argument.) Then $\phi_2(x) =$

$\phi_2(y) = u_p$. Also $x_r y_\ell = y_r x_\ell = u_{p-1}$. Thus $g(x)g(y_\ell) \in L(\phi_2(x))g(x_r)g(y_\ell) = L(u_p)g(u_{p-1})$ and $g(y)g(x_\ell) \in L(u_p)g(u_{p-1})$, which is a minimal left ideal of $f^{-1}[D_p]$ by Lemma 3.3. We have already verified that $g(x)g(y_\ell)$ and $g(y)g(x_\ell)$ are idempotents. So, since they are idempotents in the same minimal left ideal $g(x)g(y_\ell)g(y)g(x_\ell) = g(x)g(y_\ell)$, we have

$$\begin{aligned} g(x)g(y)g(x) &= g(x)g(y_\ell y)g(x_\ell x) \\ &= g(x)g(y_\ell)g(y)g(x_\ell)g(x) \\ &= g(x)g(y_\ell)g(x) \\ &= g(xx_r)g(y_\ell)g(x_\ell x) \\ &= g(x)g(x_r y_\ell x_\ell)g(x) \\ &= g(x)g(x_r x_\ell)g(x) \\ &= g(xx_r)g(x_\ell x) = g(x)g(x) = g(x). \end{aligned}$$

Consequently, $g(x)g(y)$ and $g(y)g(x)$ are idempotents. \square

Corollary 4.4. *Suppose that κ is an infinite cardinal and that each A_n or B_n is either $\{0\}$ or 2^{2^κ} . Then $D_{A,B,\omega}$ can be embedded in \mathbb{H}_κ .*

Proof. For each $p < \omega$, we give D_p the topology defined by regarding D_p as a subspace of $(\beta\kappa)^{2^{p+2}}$. We define the topology of D by taking each D_p to be clopen in D . Then D is a topological semigroup with a dense subspace of cardinality κ . The conclusion then follows from Theorems 4.3 and 2.5. \square

REFERENCES

1. J. Berglund, H. Junghenn, and P. Milnes, *Analysis on semigroups*, Wiley, N.Y., 1989. MR **91b**:43001
2. W. Comfort and S. Negrepontis, *The theory of ultrafilters*, Springer-Verlag, Berlin, 1974. MR **53**:135
3. L. Gillman and M. Jerison, *Rings of continuous functions*, van Nostrand, Princeton, 1960. MR **22**:6994
4. N. Hindman and J. Pym, *Free groups and semigroups in $\beta\mathbb{N}$* , Semigroup Forum **30** (1984), 177-193. MR **86c**:22002
5. N. Hindman and D. Strauss, *Chains of idempotents in $\beta\mathbb{N}$* , Proc. Amer. Math. Soc. **123** (1995), 3881-3888. MR **96b**:54037
6. N. Hindman and D. Strauss, *Algebra in the Stone-Čech compactification: Theory and applications*, de Gruyter, Berlin, 1998. MR **99j**:54001
7. D. McLean, *Idempotent semigroups*, Amer. Math. Monthly **61** (1954), 110-113. MR **15**:681a
8. J. Pym, *Semigroup structure in Stone-Čech compactifications*, J. London Math. Soc. **36** (1987), 421-428. MR **89b**:54043
9. W. Ruppert, *Rechsttopologische Halbgruppen*, J. Reine Angew. Math. **261** (1973), 123-133. MR **47**:6933
10. Y. Zelenyuk, *On subsemigroups of $\beta\mathbb{N}$ and absolute coretracts*, Semigroup Forum **63** (2001), 457-465. MR **2002f**:22005

DEPARTMENT OF MATHEMATICS, HOWARD UNIVERSITY, WASHINGTON, DC 20059

E-mail address: nhindman@aol.com

URL: <http://members.aol.com/nhindman/>

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF HULL, HULL HU6 7RX, UNITED KINGDOM

E-mail address: d.strauss@maths.hull.ac.uk

FACULTY OF CYBERNETICS, KYIV TARAS SHEVCHENKO UNIVERSITY, VOLODYMYRSKA STREET 64, 01033 KYIV, UKRAINE

E-mail address: grishko@i.com.ua

GALOIS GROUPS OF QUANTUM GROUP ACTIONS AND REGULARITY OF FIXED-POINT ALGEBRAS

TAKEHIKO YAMANOUCHI

Dedicated to Professor Masamichi Takesaki on the occasion of his seventieth birthday

ABSTRACT. It is shown that, for a minimal and integrable action of a locally compact quantum group on a factor, the group of automorphisms of the factor leaving the fixed-point algebra pointwise invariant is identified with the intrinsic group of the dual quantum group. It is proven also that, for such an action, the regularity of the fixed-point algebra is equivalent to the cocommutativity of the quantum group.

1. INTRODUCTION

When given an action α of a locally compact quantum group \mathbb{G} on a von Neumann algebra A , one may associate to it the subgroup $\text{Aut}(A/A^\alpha)$ of all automorphisms of A leaving the fixed-point algebra A^α invariant pointwise. Let us call this subgroup “the Galois group of α ”. As [1, Theorem III.3.3] suggests, it would sometimes happen (or be expected) that the Galois group carries an important piece of information on the quantum group \mathbb{G} itself. With this philosophy in mind, we started in [17] to investigate Galois groups of minimal actions of compact Kac algebras on factors by making good use of the Galois correspondence established by Izumi, Longo and Popa [7]. In [20], we succeeded in describing the Galois group of any minimal action of a compact Kac algebra as the so-called intrinsic group of the dual discrete Kac algebra. This extended the result of [1] cited above. As an application, we were able to show that, if the quantum group in question is finite-dimensional, then its cocommutativity is equivalent to the regularity of the fixed-point algebra. Our main goal of this paper is to extend these results to a larger class of locally compact quantum groups, not only compact Kac algebras. If we try to achieve this goal exactly along the line carried out in [17] and [20], then a Galois correspondence for a (minimal) action of a more general locally compact quantum group would certainly be needed. At the moment, it seems that the results of Enock in [4] would answer this purpose. Unfortunately, there are however a few mistakes in his proofs, and, to the best of the author’s knowledge, they have not been restored yet. So we cannot apply Enock’s Galois correspondence to the situation we will consider in this paper. Therefore, we will adopt a new approach here that does not resort to any Galois correspondence.

Received by the editors June 24, 2002 and, in revised form, November 6, 2002.

2000 *Mathematics Subject Classification.* Primary 46L65; Secondary 22D25, 46L10, 81R50.

Key words and phrases. Locally compact quantum group, action, factor, regularity.

The outline of this paper is the following. In Section 1, we fix the notation used in the whole of our discussion. Basic facts about locally compact quantum groups (in the sense of Kustermans and Vaes) and their actions on von Neumann algebras are collected. In Section 2, we will prove that the Galois group of a minimal, integrable action of a locally compact quantum group \mathbb{G} is topologically isomorphic to the intrinsic group of the dual $\widehat{\mathbb{G}}$. Section 3 is concerned with regularity of the fixed-point algebra of a minimal, integrable action. We prove that the regularity considerably determines the structure of the quantum group. Namely it is shown, with some exception, that the fixed-point algebra is regular if and only if the locally compact quantum group under consideration is cocommutative. In Section 4, we make a few remarks on the Izumi-Longo-Popa Galois correspondence. One of them concerns an explicit formula for the inverse map of their Galois correspondence. Finally, we include an Appendix for some auxiliary results which are applied to the argument made in Section 3.

The author is grateful to Professors Michel Enock and Stefaan Vaes for having informed him that there are mistakes in some proofs in [4]. He is also indebted to Professor Masaki Izumi for indicating an error in the earlier draft of the manuscript.

2. TERMINOLOGY AND NOTATION

Given a von Neumann algebra A and a faithful normal semifinite weight ϕ on A , we introduce the subsets \mathfrak{n}_ϕ , \mathfrak{m}_ϕ and \mathfrak{m}_ϕ^+ of A by

$$\mathfrak{n}_\phi = \{x \in A : \phi(x^*x) < \infty\}, \qquad \mathfrak{m}_\phi = \mathfrak{n}_\phi^* \mathfrak{n}_\phi, \qquad \mathfrak{m}_\phi^+ = \mathfrak{m}_\phi \cap A_+.$$

The standard (GNS) Hilbert space obtained from ϕ is denoted by H_ϕ . We use the symbol Λ_ϕ for the canonical embedding of \mathfrak{n}_ϕ into H_ϕ . The modular objects such as the modular operator, the modular conjugation, the S -operator, the F -operator, the modular automorphism group, etc. associated to ϕ are denoted by ∇_ϕ , J_ϕ , S_ϕ , F_ϕ , σ^ϕ , respectively. (Since we follow the notation employed in [10], the symbol ∇ will be used to denote the modular operator of a weight.) The set of unitaries in A is denoted by $\mathcal{U}(A)$. For a von Neumann subalgebra B of A , define $\mathcal{N}(B) := \{u \in \mathcal{U}(A) : uBu^* = B\}$ and call it the normalizer (group) of B in A .

We let $B(H)$ stand for the algebra of all bounded operators on a Hilbert space H .

2.1. Locally compact quantum groups.

Definition 2.1. Following [10] (see [9] also), we say that a quadruple $\mathbb{G} = (M, \Delta, \varphi, \psi)$ is a **locally compact quantum group** (in the von Neumann algebra setting) or a **von Neumann algebraic quantum group** if

- (1) M is a von Neumann algebra;
- (2) Δ is a unital normal injective $*$ -homomorphism from M into $M \otimes M$ satisfying $(\Delta \otimes id) \circ \Delta = (id \otimes \Delta) \circ \Delta$;
- (3) φ is a faithful normal semifinite weight on M such that

$$\varphi((\omega \otimes id)(\Delta(x))) = \varphi(x)\omega(1) \qquad (\forall \omega \in M_*^+, \forall x \in \mathfrak{m}_\varphi^+);$$

- (4) ψ is a faithful normal semifinite weight on M such that

$$\psi((id \otimes \omega)(\Delta(x))) = \psi(x)\omega(1) \qquad (\forall \omega \in M_*^+, \forall x \in \mathfrak{m}_\psi^+).$$

Let us fix a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ throughout the rest of this section. We will always think of M as represented on the GNS-Hilbert space H_φ obtained from φ . By the left invariance of φ , one gets a unitary $W(\mathbb{G})$ on $H_\varphi \otimes H_\varphi$ characterized by

$$W(\mathbb{G})^*(\Lambda_\varphi(x) \otimes \Lambda_\varphi(y)) = \Lambda_{\varphi \otimes \varphi}(\Delta(y)(x \otimes 1)) \quad (x, y \in \mathfrak{n}_\varphi).$$

This unitary is called the *Kac-Takesaki operator* of \mathbb{G} , and is denoted simply by W if there is no danger of confusion. The modular operator and the modular conjugation of φ will be denoted simply by ∇ and J . The *unitary antipode*, the *scaling group* and the *scaling constant* of \mathbb{G} are respectively denoted by R , $\{\tau_t\}_{t \in \mathbb{R}}$, $\nu (> 0)$. As in [10], we assume that $\psi = \varphi \circ R$.

According to [10], there canonically exists another locally compact quantum group $\widehat{\mathbb{G}} = (\widehat{M}, \widehat{\Delta}, \widehat{\varphi}, \widehat{\psi})$, called the *locally compact quantum group dual to \mathbb{G}* such that $\{\widehat{M}, H_\varphi\}$ is a standard representation. So we always regard \widehat{M} as acting on H_φ . In fact, \widehat{M} is by definition the von Neumann algebra generated by $\{(\omega \otimes id)(W) : \omega \in M_*\}$. The mapping $\lambda: \omega \in M_* \mapsto (\omega \otimes id)(W) \in \widehat{M}$ is called the *left regular representation* of \mathbb{G} . There is a canonical identification (= the Fourier transform) of $H_{\widehat{\varphi}}$ with H_φ . So we consider the modular operator and the modular conjugation of $\widehat{\varphi}$, denoted by $\widehat{\nabla}$ and \widehat{J} , as acting on H_φ . The unitary antipode, the scaling group and the modular element of $\widehat{\mathbb{G}}$ are denoted respectively by \widehat{R} , $\widehat{\tau}$, $\widehat{\delta}$. We say that \mathbb{G} is compact if $\varphi(1) < \infty$. In this case, we agree to take φ to be a state. We say that \mathbb{G} is discrete if $\widehat{\mathbb{G}}$ is compact. For the definitions of locally compact quantum groups such as the commutant \mathbb{G}' , the opposite \mathbb{G}^{op} etc., we refer the readers to [10, Section 4].

It is known that every locally compact group Γ canonically gives rise to a commutative locally compact quantum group whose underlying von Neumann algebra is $L^\infty(\Gamma)$. We denote it by $\mathbb{G}(\Gamma)$. The underlying von Neumann algebra of the dual $\widehat{\mathbb{G}}(\Gamma)$ is the group von Neumann algebra of Γ generated by the left regular representation of Γ .

We denote by $IG(\mathbb{G})$ the set of all unitaries $u \in M$ satisfying $\Delta(u) = u \otimes u$. The group $IG(\mathbb{G})$ is called the *intrinsic group* of \mathbb{G} . Next define $\mathcal{G}(\mathbb{G})$ to be the group of all automorphisms β of M such that $(\beta \otimes id) \circ \Delta = \Delta \circ \beta$. By [2] (see [19] also), $IG(\widehat{\mathbb{G}})$ is topologically isomorphic to $\mathcal{G}(\mathbb{G})$ through the mapping: $v \in IG(\widehat{\mathbb{G}}) \mapsto \beta_v := \text{Ad } v|_M \in \mathcal{G}(\mathbb{G})$. Here $IG(\widehat{\mathbb{G}})$ is endowed with the strong-operator topology, and, for a general (separable) von Neumann algebra P , we consider on the automorphism group $\text{Aut}(P)$ of P the topology of simple convergence on the predual. It is known that v is the canonical implementation of β_v .

We say that $N \subseteq M$ is a *right co-ideal* (von Neumann subalgebra) of \mathbb{G} if N is a von Neumann subalgebra of M satisfying $\Delta(N) \subseteq N \otimes M$. Thanks to [4, Théorème 3.3], we know that $N \subseteq M$ is a right co-ideal of \mathbb{G} if and only if one has

$$(2.1) \quad N = M \cap (\widehat{M} \cap N')'.$$

A (left) action of \mathbb{G} on a von Neumann algebra A is a normal injective unital $*$ -homomorphism α from A into $M \otimes A$ satisfying $(id_M \otimes \alpha) \circ \alpha = (\Delta \otimes id_M) \circ \alpha$ ([16]).

Fix an action α of \mathbb{G} on a von Neumann algebra A . By [16, Proposition 1.3], the equation

$$T_\alpha(a) := (\psi \otimes id)(\alpha(a)) \quad (a \in A_+)$$

defines a faithful normal operator-valued weight T_α from A onto $A^\alpha := \{a \in A : \alpha(a) = 1 \otimes a\}$, the fixed-point algebra A^α of α . We call T_α the operator-valued weight associated to the action α .

The crossed product of A by the action α is by definition the von Neumann algebra generated by $\alpha(A)$ and $\widehat{M} \otimes \mathbf{C}$. We denote it by $\mathbb{G}_\alpha \ltimes A$. By [16, Proposition 2.2], there exists a unique action $\hat{\alpha}$ of $\widehat{\mathbb{G}}^{op}$ on $\mathbb{G}_\alpha \ltimes A$, called the *dual action* of α , such that

$$(\mathbb{G}_\alpha \ltimes A)^{\hat{\alpha}} = \alpha(A), \qquad \hat{\alpha}(z \otimes 1) = \hat{\Delta}^{op}(z) \otimes 1 \qquad (z \in \widehat{M}).$$

For every faithful normal semifinite weight ϕ on A , by using the operator-valued weight $T_{\hat{\alpha}}$ associated to the dual action $\hat{\alpha}$, the equation

$$\tilde{\phi} := \phi \circ \alpha^{-1} \circ T_{\hat{\alpha}}$$

defines a faithful normal semifinite weight $\tilde{\phi}$ on $\mathbb{G}_\alpha \ltimes A$. The weight $\tilde{\phi}$ is called the *dual weight* of ϕ . The Hilbert space $H_{\tilde{\phi}}$ is identified with $H_\phi \otimes H_\phi$. The unitary U_ϕ on $H_\phi \otimes H_\phi$ defined by

$$U_\phi := J_{\tilde{\phi}}(\hat{J} \otimes J_\phi)$$

is called the *canonical implementation* of α (see [16]). It satisfies

$$\alpha(a) = U_\phi(1 \otimes a)U_\phi^* \qquad (a \in A).$$

By [16, Theorem 2.6], there is a unital $*$ -isomorphism Θ from the double crossed product $\widehat{\mathbb{G}}^{op} \hat{\alpha} \ltimes (\mathbb{G}_\alpha \ltimes A)$ onto $B(H_\phi) \otimes A$, and an action $\tilde{\alpha}$ of \mathbb{G} on $B(H_\phi) \otimes A$ such that

$$\begin{aligned} \tilde{\alpha} &:= \text{Ad}(\Sigma V^* \Sigma \otimes 1) \circ (\sigma \otimes \text{id}_A) \circ (\text{id}_{B(H_\phi)} \otimes \alpha), \\ (\text{Ad}(J\hat{J}) \otimes \Theta) \circ \hat{\alpha} &= \tilde{\alpha} \circ \Theta, \end{aligned}$$

where $V := (\hat{J} \otimes \hat{J})\Sigma W^*\Sigma(\hat{J} \otimes \hat{J})$ and $\Sigma: H_\phi \otimes H_\phi \rightarrow H_\phi \otimes H_\phi$ is the flip. We call $\tilde{\alpha}$ the *stabilization* of α .

We say that the action α is *integrable* if T_α is semifinite. The action α is said to be *minimal* if $A \cap (A^\alpha)' = \mathbf{C}$ and the linear span of $\{(id \otimes \omega)(\alpha(a)) : a \in A, \omega \in A_*\}$ is σ -weakly dense in M .

Finally, \mathbb{G} is called a *Kac algebra* if $\tau_t = id$ and $\sigma^\phi = \sigma^\psi$. For the general theory of Kac algebras, refer to [6].

3. REALIZATION OF INTRINSIC GROUPS IN $\text{Aut}(A/A^\alpha)$

Given a von Neumann algebra P and a von Neumann subalgebra Q of P , we define $\text{Aut}(P/Q)$ to be the group of all automorphisms of P leaving Q invariant pointwise.

Let $\mathbb{G} = (M, \Delta, \varphi, \psi)$ be a locally compact quantum group. Suppose now that we have an action α of \mathbb{G} on a von Neumann algebra A . Throughout this paper, A is always assumed to be a non-type I_n factor ($n \in \mathbf{N}$).

The mapping considered in the following proposition is essentially observed by Enock and Schwartz in [5] as a special case of their constructions of certain morphisms associated to an action of a Kac algebra. The mapping is still defined even for a general locally compact quantum group.

Proposition 3.1. *There exists a unique continuous homomorphism from $\mathcal{G}(\mathbb{G})$ into $\text{Aut}(A/A^\alpha)$ such that, with θ_β the image of $\beta \in \mathcal{G}(\mathbb{G})$ under this homomorphism, we have*

$$(\beta \otimes \text{id}) \circ \alpha = \alpha \circ \theta_\beta.$$

If the action α enjoys the property that the linear span of $\{(id_M \otimes \omega)(\alpha(a)) : a \in A, \omega \in A_\}$ is σ -weakly dense in M , then the above homomorphism is injective.*

Proof. Let $\beta \in \mathcal{G}(\mathbb{G})$ and $a \in A$. Set $X := (\beta \otimes id_A) \circ \alpha(a) \in M \otimes A$. Then

$$\begin{aligned} (id_M \otimes \alpha)(X) &= (id_M \otimes \alpha) \circ (\beta \otimes id_A) \circ \alpha(a) \\ &= (\beta \otimes id_M \otimes id_A) \circ (id_M \otimes \alpha) \circ \alpha(a) \\ &= (\beta \otimes id_M \otimes id_A) \circ (\Delta \otimes id_A) \circ \alpha(a) \\ &= (\Delta \otimes id_A) \circ (\beta \otimes id_A) \circ \alpha(a) = (\Delta \otimes id_A)(X). \end{aligned}$$

From [16, Theorem 2.7], there exists a unique element $\theta_\beta(a) \in A$ such that

$$(3.1) \quad (\beta \otimes id_A) \circ \alpha(a) = X = \alpha(\theta_\beta(a)).$$

It is easy to check by using (3.1) that θ_β is a homomorphism from A into itself. Moreover, one can easily verify that $\theta_{\beta_1 \beta_2} = \theta_{\beta_1} \circ \theta_{\beta_2}$ and $\theta_{id} = id$. Hence θ_β is an automorphism of A . That the restriction of θ_β to A^α is the identity follows also from (3.1). Thus the mapping $\beta \in \mathcal{G}(\mathbb{G}) \mapsto \theta_\beta \in \text{Aut}(A/A^\alpha)$ is indeed a homomorphism. Because of (3.1), we find that $(\beta \otimes id_A)|_{\alpha(A)}$ is an automorphism. With this in mind, θ_β has the form

$$\theta_\beta = \alpha^{-1} \circ (\beta \otimes id_A)|_{\alpha(A)} \circ \alpha.$$

Hence $\beta \mapsto \theta_\beta$ is continuous.

Now suppose that the linear span of $\{(id_M \otimes \omega)(\alpha(a)) : a \in A, \omega \in A_*\}$ is σ -weakly dense in M . If $\theta_\beta = id$, then (3.1) implies that β is the identity on $\{(id_M \otimes \omega)(\alpha(a)) : a \in A, \omega \in A_*\}$. So $\beta = id$. Consequently, the homomorphism in question is injective. \square

Lemma 3.2. *Let β be an automorphism of M such that there is a $\theta \in \text{Aut}(A)$ such that $(\beta \otimes id) \circ \alpha = \alpha \circ \theta$. Suppose that the linear span of $\{(id_M \otimes \omega)(\alpha(a)) : a \in A, \omega \in A_*\}$ is σ -weakly dense in M . Then β belongs to $\mathcal{G}(\mathbb{G})$, and one has $\theta = \theta_\beta$.*

Proof. Note first that, if θ is an automorphism as above, then it automatically belongs to $\text{Aut}(A/A^\alpha)$. Therefore, it suffices by Proposition 3.1 to show that β belongs to $\mathcal{G}(\mathbb{G})$.

Let $a \in A$ and $\omega \in A_*$. Then we have

$$\begin{aligned} &(\beta \otimes id_M) \circ \Delta((id_M \otimes \omega)(\alpha(a))) \\ &= (\beta \otimes id_M) \circ (id_M \otimes id_M \otimes \omega) \circ (\Delta \otimes id_A)(\alpha(a)) \\ &= (id_M \otimes id_M \otimes \omega) \circ (\beta \otimes id_M \otimes id_A) \circ (id_M \otimes \alpha)(\alpha(a)) \\ &= (id_M \otimes id_M \otimes \omega) \circ (id_M \otimes \alpha) \circ (\beta \otimes id_A)(\alpha(a)) \\ &= (id_M \otimes id_M \otimes \omega) \circ (id_M \otimes \alpha) \circ \alpha(\theta(a)) \\ &= (id_M \otimes id_M \otimes \omega) \circ (\Delta \otimes id_A) \circ \alpha(\theta(a)) \\ &= \Delta \circ \beta((id_M \otimes \omega)(\alpha(a))). \end{aligned}$$

From our assumption on α , it follows that β satisfies $(\beta \otimes id) \circ \Delta = \Delta \circ \beta$. \square

Let ι be the trivial action of \mathbb{G} on \mathbf{C} . Namely, ι is the mapping from \mathbf{C} into $M \otimes \mathbf{C}$ defined by $\iota(c) := 1 \otimes c$ ($c \in \mathbf{C}$). Then the crossed product $\mathbb{G}_\iota \ltimes \mathbf{C}$ is (canonically isomorphic to) \widehat{M} , and the dual action $\hat{\iota}$ is the coproduct $\hat{\Delta}^{op}$. It is also clear that the dual weight of $t_{\mathbf{C}}$ corresponds to $\hat{\varphi}$. In this case, the stabilization $\tilde{\iota}$ of ι has the form $\tilde{\iota}(x) = \Sigma V^* \Sigma (1 \otimes x) \Sigma V \Sigma$ for any $x \in B(H_\varphi)$. Therefore, we obtain

$$(3.2) \quad \tilde{\alpha}(x \otimes 1) = \tilde{\iota}(x) \otimes 1$$

for any $x \in B(H_\varphi)$.

Lemma 3.3. *Let z be in $B(H_\varphi)$ such that $z \otimes 1 \in \mathbb{G}_\alpha \ltimes A$. Then z belongs to \widehat{M} .*

Proof. Let z be an element as above. Since $(B(H_\varphi) \otimes A)^{\tilde{\alpha}} = \mathbb{G}_\alpha \ltimes A$, we have $\tilde{\alpha}(z \otimes 1) = 1 \otimes z \otimes 1$. On the other hand, by (3.2), we have $\tilde{\alpha}(z \otimes 1) = \tilde{\iota}(z) \otimes 1$. Hence we obtain $\tilde{\iota}(z) = 1 \otimes z$. Since $B(H_\varphi)^{\tilde{\iota}} = \widehat{M}$, z must belong to \widehat{M} . \square

Lemma 3.4. *Suppose that α is minimal, Then we have $\alpha(A)' \cap B(H_\varphi) \otimes A = M' \otimes \mathbf{C}$.*

Proof. It is clear that $M' \otimes \mathbf{C}$ is contained in $\alpha(A)' \cap B(H_\varphi) \otimes A$. Take any $T \in \alpha(A)' \cap B(H_\varphi) \otimes A$. Since T particularly commutes with any element of $\alpha(A^\alpha) = \mathbf{C} \otimes A^\alpha$, it follows from the minimality of α that T belongs to $B(H_\varphi) \otimes \mathbf{C}$. So it has the form $T = y \otimes 1$ for some $y \in B(H_\varphi)$. If $a \in A$ and $\omega \in A_*$, then we have

$$y(id \otimes \omega)(\alpha(a)) = (id \otimes \omega)(T\alpha(a)) = (id \otimes \omega)(\alpha(a)T) = (id \otimes \omega)(\alpha(a))y.$$

By minimality of α , y must be in M' . \square

Since $(\mathbb{G}_\alpha \ltimes A)^{\hat{\alpha}} = \alpha(A)$, it follows from Proposition 3.1 that there exists a continuous homomorphism $\beta \in \mathcal{G}(\widehat{\mathbb{G}}^{op}) \mapsto \theta_\beta \in \text{Aut}(\mathbb{G}_\alpha \ltimes A/\alpha(A))$ satisfying $(\beta \otimes id) \circ \hat{\alpha} = \hat{\alpha} \circ \theta_\beta$. Since $\hat{\alpha}$ enjoys the property mentioned in Proposition 3.1, the homomorphism $\beta \mapsto \theta_\beta$ is injective in this case.

Lemma 3.5. *If the action α is minimal, then the homomorphism $\beta \in \mathcal{G}(\widehat{\mathbb{G}}^{op}) \mapsto \theta_\beta \in \text{Aut}(\mathbb{G}_\alpha \ltimes A/\alpha(A))$ is a topological isomorphism.*

Proof. Let θ be in $\text{Aut}(\mathbb{G}_\alpha \ltimes A/\alpha(A))$. If $z \in \widehat{M}$ and $b \in A^\alpha$, then

$$\begin{aligned} \theta(z \otimes 1)(1 \otimes b) &= \theta(z \otimes 1)\alpha(b) = \theta((z \otimes 1)\alpha(b)) = \theta(z \otimes b) \\ &= \theta(\alpha(b)(z \otimes 1)) = (1 \otimes b)\theta(z \otimes 1). \end{aligned}$$

This shows that $\theta(\widehat{M} \otimes \mathbf{C}) \subseteq (\mathbf{C} \otimes A^\alpha)' = B(H_\varphi) \otimes (A^\alpha)'$. From this, we obtain

$$\theta(\widehat{M} \otimes \mathbf{C}) \subseteq B(H_\varphi) \otimes \{A \cap (A^\alpha)'\} = B(H_\varphi) \otimes \mathbf{C}.$$

Hence, for any $z \in \widehat{M}$, there is a unique $\beta_\theta(z) \in B(H_\varphi)$ such that

$$(3.3) \quad \theta(z \otimes 1) = \beta_\theta(z) \otimes 1.$$

Thanks to Lemma 3.3, $\beta_\theta(z)$ belongs to \widehat{M} . Due to (3.3), it is easy to see that β_θ is an automorphism of \widehat{M} , and that $\beta_{\theta_1\theta_2} = \beta_{\theta_1} \circ \beta_{\theta_2}$, $\beta_{id} = id$.

If $a \in A$, then

$$(3.4) \quad (\beta_\theta \otimes id)(\hat{\alpha}(\alpha(a))) = (\beta_\theta \otimes id)(1 \otimes \alpha(a)) = 1 \otimes \alpha(a) = \hat{\alpha}(\theta(\alpha(a))).$$

Fix a faithful normal semifinite weight ω on A , and regard A as represented on H_ω . Let \tilde{J}_ω be the modular conjugation of the dual weight $\tilde{\omega}$ and U the canonical implementation of α on $H_\varphi \otimes H_\omega$. So $U = \tilde{J}_\omega(J \otimes J_\omega)$. Choose the canonical implementation unitary V of θ on $H_\varphi \otimes H_\omega$. Then we have $V \in \alpha(A)'$. Since V commutes with \tilde{J}_ω , we also have $V \in \tilde{J}_\omega \alpha(A)' \tilde{J}_\omega = (\mathbf{C} \otimes J_\omega A J_\omega)' = B(H_\varphi) \otimes A$. It follows from Lemma 3.4 that there exists a unitary $v \in M'$ such that $V = v \otimes 1$. Therefore we have $\beta_\theta(z) = v z v^*$ for any $z \in \widehat{M}$. By [12, Proposition I.9], which is valid also for any locally compact quantum group, we see that rv belongs to $IG(\mathbb{G}')$ for some $r \in \mathbf{C}$ with $|r| = 1$. Since both V and rV are the canonical implementation of θ , we must have $r = 1$. So v is in $IG(\mathbb{G}')$. It follows that β_θ belongs to $\mathcal{G}(\widehat{\mathbb{G}}^{op})$. Let $z \in \widehat{M}$. Then we have

$$\begin{aligned} \hat{\alpha}(\theta(z \otimes 1)) &= \hat{\alpha}(\beta_\theta(z) \otimes 1) = \hat{\Delta}^{op}(\beta_\theta(z)) \otimes 1 \\ &= (\beta_\theta \otimes id) \circ \hat{\Delta}^{op}(z) \otimes 1 = (\beta_\theta \otimes id) \circ \hat{\alpha}(z \otimes 1). \end{aligned}$$

From this, together with (3.4), we get $(\beta_\theta \otimes id) \circ \hat{\alpha} = \hat{\alpha} \circ \theta$. By Lemma 3.2, we find that $\theta = \theta_{\beta_\theta}$. Thus we have shown the surjectivity. The inverse map is also continuous due to (3.3). \square

Theorem 3.6. *If α is a minimal, integrable action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A , then there exists a topological isomorphism $\beta \in \mathcal{G}(\mathbb{G}) \mapsto \theta_\beta \in \text{Aut}(A/A^\alpha)$ with the property $(\beta \otimes id) \circ \alpha = \alpha \circ \theta_\beta$.*

Proof. If A^α is infinite, then, by [16, Proposition 6.4], α is a dual action. Hence the assertion follows from Lemma 3.5.

To deal with a general case, take a (separable) infinite factor L , and put $\bar{A} := L \otimes A$. Also define $\bar{\alpha} := (\sigma \otimes id) \circ (id_L \otimes \alpha)$, which is an action of \mathbb{G} on \bar{A} with $\bar{A}^{\bar{\alpha}} = L \otimes A^\alpha$ infinite. Remark that $\text{Aut}(\bar{A}/\bar{A}^{\bar{\alpha}}) = \{id_L \otimes \theta : \theta \in \text{Aut}(A/A^\alpha)\}$. Let $\theta \in \text{Aut}(A/A^\alpha)$. By the previous paragraph and the above remark, there exists $\beta \in \mathcal{G}(\mathbb{G})$ such that $(\beta \otimes id) \circ \bar{\alpha} = \bar{\alpha} \circ (id_L \otimes \theta)$. But this yields $(\beta \otimes id) \circ \alpha = \alpha \circ \theta$. \square

4. REGULARITY OF A^α IN A

As in the previous section, let α be a minimal action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A .

We represent A on a (separable) Hilbert space K so that $\{A, K, J_A\}$ is a standard representation. Let $u \in \mathcal{N}(A^\alpha)$. Then the restriction of $\text{Ad } u$ to $(A^\alpha)'$ clearly defines an automorphism θ_u in $\text{Aut}((A^\alpha)'/A')$. The homomorphism $u \in \mathcal{N}(A^\alpha) \mapsto \theta_u \in \text{Aut}((A^\alpha)'/A')$ obviously has $\mathcal{U}(A^\alpha)$ as its kernel. The basic extension for $A^\alpha \subseteq A$ is denoted by A_1 . So we have $A_1 = J_A(A^\alpha)'J_A$. If A^α is infinite, then so is A . Thanks to [3, Corollaire 1], we may then choose the above Hilbert space K in such a way that there is a unit vector $\xi_0 \in K$ that is cyclic and separating for both A and A^α . Let J_A then denote the modular conjugation of A associated to ξ_0 .

Lemma 4.1. *Suppose as above that A^α is infinite. Then the homomorphism $u \in \mathcal{N}(A^\alpha) \mapsto \theta_u \in \text{Aut}((A^\alpha)'/A')$ defined above is surjective, that is, it has $\text{Aut}((A^\alpha)'/A')$ as its image.*

Proof. Let θ be in $\text{Aut}((A^\alpha)'/A')$. Since $\{(A^\alpha)', K\}$ is a standard representation, there exists a unitary v on K such that $\theta = \text{Ad } v|_{(A^\alpha)'}$. Since $\theta|_{A'} = id$, v belongs to $(A')' = A$. It is easy to see that v is in $\mathcal{N}(A^\alpha)$. We clearly have $\theta = \theta_v$. \square

Remark. If A is finite and \mathbb{G} is finite-dimensional (so that the Jones index $[A : A^\alpha]$ is finite), then it follows from [11, Proposition 1.7] that the map $u \in \mathcal{N}(A^\alpha) \mapsto \theta_u \in \text{Aut}((A^\alpha)' / A')$ is surjective.

From now on, we assume that α is *minimal* and *integrable* (A is not necessarily infinite). Fix a faithful normal semifinite weight ω on A , and represent A on H_ω now. Let U be the canonical implementation of α associated to ω . Due to [16, Proposition 6.2], this assumption is equivalent to the one that α is *outer* and integrable. From [16, Corollary 5.6] and [16, Proposition 6.2], it follows that the inclusions

$$\mathbf{C} \otimes A^\alpha \subseteq \alpha(A) \subseteq \mathbb{G}_{\alpha \ltimes} A \quad \text{and} \quad A^\alpha \subseteq A \subseteq A_1$$

are isomorphic. According to [16, Corollary 5.6], the isomorphism $\rho: \mathbb{G}_{\alpha \ltimes} A \rightarrow A_1$ is characterized by

$$(4.1) \quad \rho(\alpha(a)) = a \quad (a \in A),$$

$$(4.2) \quad \rho((\phi \otimes id)(W) \otimes 1) = (\phi \otimes id)(U^*) \quad (\phi \in M_*).$$

Incidentally, Equation (4.2) can be rewritten as

$$(4.3) \quad \rho(\lambda'(\phi) \otimes 1) = (\hat{J}J\phi J\hat{J} \otimes id)(U) \quad (\phi \in (M')_*),$$

where λ' stands for the left regular representation of \mathbb{G}' that is given by $\lambda'(\phi) = \lambda(\hat{J}J\phi^*J\hat{J})^*$. We will make use of this isomorphism in the discussion that follows. Since $A_1 = J_\omega(A^\alpha)'J_\omega$ and $A = J_\omega A'J_\omega$, $\text{Aut}((A^\alpha)' / A')$ is isomorphic to $\text{Aut}(\mathbb{G}_{\alpha \ltimes} A / \alpha(A))$. Thus we obtain a homomorphism from $\mathcal{N}(A^\alpha)$ into $\text{Aut}(\mathbb{G}_{\alpha \ltimes} A / \alpha(A))$. By using the isomorphism ρ introduced above, it is explicitly given as follows: $u \in \mathcal{N}(A^\alpha) \mapsto \theta_u := \rho^{-1} \circ \text{Ad}(J_\omega u J_\omega)|_{A_1} \circ \rho$. As we saw in Lemma 4.1 and the remark after it, this homomorphism is surjective if A^α is infinite or if A is finite and \mathbb{G} is finite-dimensional. But it may not be in general. So our next goal is to identify its image in detail. For this, first note that, thanks to Proposition 3.5, it is enough to identify automorphisms $\beta \in \mathcal{G}(\widehat{\mathbb{G}}^{op})$ such that $\theta_\beta = \theta_u$ for $u \in \mathcal{N}(A^\alpha)$. Moreover, since each θ_β has the form $\theta_\beta = \text{Ad}(v \otimes 1)$ for a unique $v \in IG(\mathbb{G}')$ with $\beta = \text{Ad} v$, it suffices to identify unitaries $v \in IG(\mathbb{G}')$ such that $\text{Ad}(v \otimes 1) = \theta_u$ for some $u \in \mathcal{N}(A^\alpha)$.

Proposition 4.2. *Let α be a minimal and integrable action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A .*

- (1) *For any $u \in \mathcal{N}(A^\alpha)$, there exists a unique unitary $w(u) \in IG(\mathbb{G})$ such that $\alpha(u)(1 \otimes u^*) = w(u) \otimes 1$.*
- (2) *For any $u \in \mathcal{N}(A^\alpha)$, with $w(u)$ in Part (1), we have*

$$\theta_u = \text{Ad}(Jw(u)J \otimes 1) = \text{Ad}(J \otimes J_\omega)\alpha(u)(1 \otimes u^*)(J \otimes J_\omega).$$

We denote the unitary $Jw(u)J$ in $IG(\mathbb{G}')$ by $v(u)$.

Proof. Let $u \in \mathcal{N}(A^\alpha)$.

(1) It is straightforward to check that $\alpha(u)(1 \otimes u^*)$ commutes with any element of the form $1 \otimes a$, where $a \in A^\alpha$. So it is contained in $M \otimes A \cap (\mathbf{C} \otimes A^\alpha)' = M \otimes \mathbf{C}$. Hence there exists a unitary $w \in M$ such that $\alpha(u)(1 \otimes u^*) = w \otimes 1$. Thus $\alpha(u) = w \otimes u$. By applying $\Delta \otimes id$ to both sides of this identity, we obtain $\Delta(w) = w \otimes w$. Therefore w belongs to $IG(\mathbb{G})$. So put $w(u) := w$.

(2) Choose $\beta \in \mathcal{G}(\widehat{\mathbb{G}}^{op})$ and $v \in IG(\mathbb{G}')$ such that $\theta_u = \theta_\beta$ and $\beta = \text{Ad } v$. Since $\widehat{\mathbb{G}}' = \widehat{\mathbb{G}}^{op}$, we have $\beta(\lambda'(\phi)) = \lambda'(v\phi)$. Therefore, by (4.3), the identity $\theta_u = \theta_\beta$ is equivalent to the next:

$$\text{Ad}(J_\omega u J_\omega)(\hat{J} J \phi J \hat{J} \otimes id)(U) = (\hat{J} J v \phi J \hat{J} \otimes id)(U) \quad (\phi \in (M')_*).$$

This is further equivalent to

$$(J \hat{J} \otimes J_\omega u J_\omega) U (J \hat{J} \otimes J_\omega u^* J_\omega) = (J \hat{J} \otimes 1) U (J \hat{J} v \otimes 1).$$

By using $U^* = (\hat{J} \otimes J_\omega) U (\hat{J} \otimes J_\omega)$, we can reduce the above identity to

$$\alpha(u)(1 \otimes u^*) = J v J \otimes 1.$$

Consequently, we obtain $w(u) = J v J$. □

Definition 4.3. Let β be an action of \mathbb{G} on a von Neumann algebra P . For a right co-ideal N of \mathbb{G} , the intermediate von Neumann subalgebra $P(N)$ of $P^\beta \subseteq P$ associated to N (see [7]) is defined by

$$P(N) := \{x \in P : \beta(x) \in N \otimes P\}.$$

The following lemma is proven in [4] for the case where \mathbb{G} is a Woronowicz algebra. The claim and its proof are still valid even for a general locally compact quantum group.

Lemma 4.4 (Proposition 3.5, [4]). *Let N be a right co-ideal of \mathbb{G} . Then we have*

$$N = \{(id \otimes \omega)(\Delta(x)) : \omega \in M_*, a \in N\}''.$$

Proof. Denote by N_1 the right-hand side of the above claim. Clearly we have $N_1 \subseteq N$. Let $x \in N$. For any $y \in N'_1$, $\rho \in B(H_\varphi)_*$ and $\omega \in M_*$, we have

$$\begin{aligned} (\rho \otimes \omega)((y \otimes 1)\Delta(x)) &= \rho(y(id \otimes \omega)(\Delta(x))) \\ &= \rho((id \otimes \omega)(\Delta(x))y) = (\rho \otimes \omega)(\Delta(x)(y \otimes 1)). \end{aligned}$$

This shows that $\Delta(x)$ is in $N_1 \otimes M$. Hence $\Delta(N) \subseteq N_1 \otimes M$. In particular, N_1 is also a right co-ideal of \mathbb{G} . Therefore, $\gamma := \Delta^{op}|_{N_1}$ defines an action γ of \mathbb{G}^{op} on N_1 . If $x \in N$, then, by the above result, $\Delta^{op}(x) \in M \otimes N_1$. Moreover, we have

$$(id \otimes \gamma)(\Delta^{op}(x)) = (id \otimes \Delta^{op})(\Delta^{op}(x)) = (\Delta^{op} \otimes id)(\Delta^{op}(x)).$$

From [16, Theorem 2.7], it follows that $\Delta^{op}(x)$ belongs to $\gamma(N_1) = \Delta^{op}(N_1)$. Thus $x \in N_1$. □

Lemma 4.5. *Let β be an action of \mathbb{G} on a von Neumann algebra P . For any intermediate von Neumann subalgebra Q of $P^\alpha \subseteq P$, $\{(id \otimes \omega)(\beta(x)) : x \in Q, \omega \in P_*\}''$ is a right co-ideal of \mathbb{G} . We denote this right co-ideal by $N_\beta(Q)$, and call it the right co-ideal associated to Q .*

Proof. Let $y \in N_\beta(Q)'$. For any $\rho, \phi \in B(H_\varphi)_*$, $\omega \in P_*$ and $x \in Q$, we have

$$\begin{aligned} &\langle (y \otimes 1)\Delta((id \otimes \omega)(\beta(x))), \rho \otimes \phi \rangle \\ &= \langle (y \otimes 1)(id \otimes id \otimes \omega)((id \otimes \beta)(\beta(x))), \rho \otimes \phi \rangle \\ &= \langle y(id \otimes (\phi \otimes \omega) \circ \beta)(\beta(x)), \rho \rangle \\ &= \langle (id \otimes (\phi \otimes \omega) \circ \beta)(\beta(x))y, \rho \rangle \\ &= \langle \Delta((id \otimes \omega)(\beta(x)))(y \otimes 1), \rho \otimes \phi \rangle. \end{aligned}$$

From this, we see that $\Delta((id \otimes \omega)(\beta(x)))$ is included in $N_\beta(Q) \otimes M$. Hence $N_\beta(Q)$ is a right co-ideal of \mathbb{G} . \square

Lemma 4.6. *Let β be an action of \mathbb{G} on a von Neumann algebra P . With the notation introduced above, we have $N = N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N))$ for any right co-ideal N of $\widehat{\mathbb{G}}^{op}$. If L is the von Neumann subalgebra of $\mathbb{G}_\beta \ltimes P$ generated by $\beta(P)$ and $N \otimes \mathbb{C}$ for some right co-ideal N of $\widehat{\mathbb{G}}^{op}$, then $N_{\hat{\beta}}(L) = N$.*

Proof. Let N be a right co-ideal of $\widehat{\mathbb{G}}^{op}$. It is plain to see that $N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N))$ is contained in N . Let $x \in N$ and $\rho \in B(H_\varphi)_*$. Since $\hat{\Delta}^{op}(x) \in N \otimes \widehat{M}$, we have

$$\hat{\beta}(x \otimes 1) = \hat{\Delta}^{op}(x) \otimes 1 \in N \otimes (\mathbb{G}_\beta \ltimes P).$$

Thus $x \otimes 1$ belongs to $(\mathbb{G}_\beta \ltimes P)(N)$. If $\omega \in P_*$ is a state, then

$$\begin{aligned} (id \otimes \rho)(\hat{\Delta}^{op}(x)) &= (id \otimes \rho \otimes \omega)(\hat{\Delta}^{op}(x) \otimes 1) \\ &= (id \otimes \rho \otimes \omega)(\hat{\beta}(x \otimes 1)) \in N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N)). \end{aligned}$$

It follows from Lemma 4.4 that N is contained in $N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N))$. Therefore, we have proven that $N = N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N))$.

Let L be as above. Take any state $\phi \in A_*$. Then, by Lemma 4.4, we have

$$\begin{aligned} N &= \{(id \otimes \omega)(\hat{\Delta}^{op}(x)) : x \in N, \omega \in \widehat{M}_*\}'' \\ &= \{(id \otimes \omega \otimes \phi)(\hat{\Delta}^{op}(x) \otimes 1) : x \in N, \omega \in \widehat{M}_*\}'' \\ &= \{(id \otimes \omega \otimes \phi)(\hat{\beta}(x \otimes 1)) : x \in N, \omega \in \widehat{M}_*\}'' \subseteq N_{\hat{\beta}}(L). \end{aligned}$$

In the meantime, L is clearly included in $(\mathbb{G}_\beta \ltimes P)(N)$. Hence, by the result of the previous paragraph, we get $N_{\hat{\beta}}(L) \subseteq N_{\hat{\beta}}((\mathbb{G}_\beta \ltimes P)(N)) = N$. \square

Lemma 4.7. *Let $\mathbb{G} = (M, \Delta, \varphi, \psi)$ be a locally compact quantum group. Then $M^{\mathcal{G}(\mathbb{G})} = \mathbb{C}$ if and only if \mathbb{G} is commutative.*

Proof. If \mathbb{G} is commutative, then we clearly have $M^{\mathcal{G}(\mathbb{G})} = \mathbb{C}$.

Put $N := IG(\widehat{\mathbb{G}})''$. It is easy to see that N is a two-sided co-ideal (von Neumann subalgebra) of \widehat{M} (more precisely, of $\widehat{\mathbb{G}}$). Moreover, we have $M^{\mathcal{G}(\mathbb{G})} = M \cap N'$. From [4, Théorème 3.3] (which is still valid for a locally compact quantum group), it follows that $N = \widehat{M}$ if $M^{\mathcal{G}(\mathbb{G})} = \mathbb{C}$. Then $\widehat{\mathbb{G}}$ is cocommutative. \square

As explained in Section 1, the mapping $v \in IG(\mathbb{G}') \mapsto \text{Ad } v|_{\widehat{M}} \in \mathcal{G}(\widehat{\mathbb{G}}^{op})$ is a topological isomorphism. Let $\beta_v \in \mathcal{G}(\widehat{\mathbb{G}}^{op})$ stand for the image of $v \in IG(\mathbb{G}')$ under this isomorphism.

Theorem 4.8. *Let α be a minimal and integrable action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A . We set $P := \mathcal{N}(A^\alpha)''$ and define P_1 to be the basic extension of $P \subseteq A$, i.e, $P_1 := J_\omega P' J_\omega$. With the isomorphism $\rho: \mathbb{G}_\alpha \ltimes A \rightarrow A_1$, put $Q_1 := \rho^{-1}(P_1)$. Let $w(\cdot): \mathcal{N}(A^\alpha) \rightarrow IG(\mathbb{G})$ and $v(\cdot): \mathcal{N}(A^\alpha) \rightarrow IG(\mathbb{G}')$ be the maps obtained in Proposition 4.2.*

- (1) *The maps $w(\cdot)$, $v(\cdot)$ are continuous homomorphisms with $\mathcal{U}(A^\alpha)$ their kernel.*

(2) We have

$$\begin{aligned} Q_1 &= (\mathbb{G} \ltimes A)^{\{\theta_u: u \in \mathcal{N}(A^\alpha)\}} = (\mathbb{G} \ltimes A)^{\{\theta_{\beta_{v(u)}}: u \in \mathcal{N}(A^\alpha)\}} \\ &= (\mathbb{G} \ltimes A)(\widehat{M}^{\{\beta_{v(u)}: u \in \mathcal{N}(A^\alpha)\}}). \end{aligned}$$

(3) We have $N_{\hat{\alpha}}(Q_1) = \widehat{M}^{\{\beta_{v(u)}: u \in \mathcal{N}(A^\alpha)\}}$.

Proof. (1) This is straightforward.

(2) With the original definition of θ_u ($u \in \mathcal{N}(A^\alpha)$), we clearly have that $(A_1)^{\{\theta_u: u \in \mathcal{N}(A^\alpha)\}} = P_1$. So we get the first equality of our assertion. By Proposition 4.2 (2), we have $\theta_u = \beta_{v(u)}$, which yields the second equality. The last equality follows from the fact that θ_β always satisfies $(\beta \otimes id) \circ \hat{\alpha} = \hat{\alpha} \circ \theta_\beta$ due to Proposition 3.1.

(3) This follows from Part (2) and Lemma 4.6. \square

Let $P \subseteq Q$ be an inclusion of von Neumann algebras. If the normalizer group $\mathcal{N}(P)$ of P in Q generates Q , we say that P is *regular* in Q .

The next theorem is a direct generalization of [20, Theorem 3.6], where we treated only the case where \mathbb{G} is finite-dimensional.

Theorem 4.9. *Suppose that α is a minimal and integrable action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A .*

(1) *If A^α is regular in A , then \mathbb{G} is cocommutative.*

(2) *If A^α is infinite, or if A is finite and \mathbb{G} is finite-dimensional, then the cocommutativity of \mathbb{G} implies the regularity of A^α in A .*

Proof. Retain the notation employed in Theorem 4.8.

(1) Suppose that A^α is regular in A . Then $Q_1 = \alpha(A)$. By Theorem 4.8 (3), we get

$$\widehat{M} \cap \{v(u) : u \in \mathcal{N}(A^\alpha)\}' = \widehat{M}^{\{\beta_{v(u)}: u \in \mathcal{N}(A^\alpha)\}} = \mathbf{C}.$$

In particular, $\widehat{M} \cap \{w(u) : u \in \mathcal{N}(A^\alpha)\}' = \mathbf{C}$, because we have $J\widehat{M}J = \widehat{M}$ in general (see [10]). Since $\{w(u) : u \in \mathcal{N}(A^\alpha)\}''$ is a two-sided co-ideal of \mathbb{G} , it follows from (2.1) that

$$\{w(u) : u \in \mathcal{N}(A^\alpha)\}'' = M \cap (\widehat{M} \cap \{w(u) : u \in \mathcal{N}(A^\alpha)\})' = M.$$

Hence \mathbb{G} is cocommutative.

(2) Assume that A^α is infinite, or that A is finite and \mathbb{G} is finite-dimensional. Let $v \in IG(\mathbb{G}')$. It follows from the proof of Lemma 3.5 that $\text{Ad}(v \otimes 1)$ is in $\text{Aut}(\mathbb{G} \ltimes A / \alpha(A))$. By assumption, it also follows from Proposition 4.2 and the discussion preceding it that $u \in \mathcal{N}(A^\alpha) \mapsto \theta_u = \text{Ad}(v(u) \otimes 1) \in \text{Aut}(\mathbb{G} \ltimes A / \alpha(A))$ is an isomorphism. Hence there is a unique $u \in \mathcal{N}(A^\alpha)$ such that

$$\text{Ad}(v \otimes 1) = \theta_u = \text{Ad}(v(u) \otimes 1).$$

By the proof of Lemma 3.5 again, we must have $v = v(u)$. Therefore, the map $v(\cdot)$ is surjective, i.e., $v(\mathcal{N}(A^\alpha)) = IG(\mathbb{G}')$. So, if \mathbb{G} is cocommutative, then $v(\mathcal{N}(A^\alpha))' = M$. From this, it follows that

$$\widehat{M}^{\{\beta_{v(u)}: u \in \mathcal{N}(A^\alpha)\}} = \widehat{M} \cap M = \mathbf{C}.$$

By Theorem 4.8 (3), $N_{\hat{\alpha}}(Q_1) = \mathbf{C}$. This means that $Q_1 = (\mathbb{G} \ltimes A)^{\hat{\alpha}} = \alpha(A)$. Therefore $A = P$. \square

Next we would like to discuss Theorem 4.9 (2) in the case where A is finite and \mathbb{G} is infinite-dimensional.

Let Γ be a (countable) discrete group and α be a minimal co-action of Γ (i.e., a minimal action of $\widehat{\mathbb{G}}(\Gamma)$) on a type II_1 factor A . For any $\gamma \in \Gamma$, define

$$A^\alpha(\gamma) := \{a \in A : \alpha(a) = \lambda(\gamma) \otimes a\}$$

and call it the eigensubspace of γ . The subspaces $\{A^\alpha(\gamma)\}_{\gamma \in \Gamma}$ played a vital part in defining the Connes spectrum $\Gamma(\alpha)$ of α in [18].

Proposition 4.10. *Let Γ , α and A be as above. For any $\gamma \in \Gamma$, the eigensubspace $A^\alpha(\gamma)$ contains a unitary.*

Proof. By [18, Theorem 3.17], $A^\alpha(\gamma)^* A^\alpha(\gamma)$ is σ -weakly dense in A^α , so that $A^\alpha(\gamma)$ contains plenty of nonzero elements for any $\gamma \in \Gamma$. Fix an arbitrary $\gamma \in \Gamma \setminus \{e\}$. Put

$$B := \left\{ \begin{bmatrix} a & X \\ Y^* & b \end{bmatrix} : a, b \in A^\alpha, X, Y \in A^\alpha(\gamma) \right\}.$$

By using the minimality of α and the fact that $A^\alpha \cap A^\alpha(\gamma) = \{0\}$ ($\forall \gamma \neq e$), one can easily verify that B is a subfactor of $M_2(A) = A \otimes M_2(\mathbf{C})$. Accordingly, the unique tracial state on $A \otimes M_2(\mathbf{C})$ restricts to that of B . So the projections

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

are equivalent in B , since they have equal traces there. Hence there exists an isometry $Y \in A^\alpha(\gamma)$. Since A is finite, Y is a unitary. \square

Theorem 4.11. *Suppose that α is a minimal, integrable action of a cocommutative locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a II_1 factor A . Then A^α is regular in A .*

Proof. Since A is finite, it follows that \mathbb{G} must be of compact type. Hence \mathbb{G} has the form $\mathbb{G} = \widehat{\mathbb{G}}(\Gamma)$ for a unique (countable) discrete group Γ . By Proposition 4.10, every eigensubspace $A^\alpha(\gamma)$ contains a unitary $V(\gamma)$. Clearly $V(\gamma)$ belongs to $\mathcal{N}(A^\alpha)$. So it remains to show that $\{V(\gamma)\}_{\gamma \in \Gamma}$ and A^α together generate A . But this follows from the next two facts: (i) A is generated by $\{A^\alpha(\gamma)\}_{\gamma \in \Gamma}$; (ii) $A^\alpha(\gamma)$ has the form $A^\alpha(\gamma) = A^\alpha V(\gamma)$ for any $\gamma \in \Gamma$. \square

Remark. We remark that, for a minimal, integrable action α on an infinite factor A with A^α finite, the cocommutativity of \mathbb{G} does NOT in general imply the regularity of A^α . In fact, suppose that A is a factor of type III_λ ($0 < \lambda < 1$). Take a faithful normal state ω on A with $\sigma_T^\omega = \text{id}_A$, where $T := -2\pi/\log \lambda$. We regard this modular action as an action α of the one-dimensional torus \mathbf{T} on A . It is well known that α is a minimal (integrable) action. Note that A^α , the centralizer of ω , is a factor of type II_1 . Let u be in $\mathcal{N}(A^\alpha)$. It is easy to see that $u^* \alpha_z(u)$ lies in $(A^\alpha)' \cap A = \mathbf{C}$. It follows that u belongs to some spectral subspace $A^\alpha(n) := \{a \in A : \alpha_z(a) = z^n a \ (\forall z \in \mathbf{T})\}$ ($n \in \mathbf{Z}$) of the action α . But, according to [14, Lemma 1.6], every spectral subspace $A^\alpha(n)$ except $A^\alpha(0) = A^\alpha$ contains no unitary. This shows that $\mathcal{N}(A^\alpha)$ is contained in A^α . Therefore A^α is not regular in A .

5. REMARKS ON ILP'S GALOIS CORRESPONDENCE

Let α be an action of a compact Kac algebra $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A . In what follows, the action α is assumed to be minimal and integrable, but we do not necessarily assume that A^α is infinite.

In [7], a complete Galois correspondence for a minimal action of a compact Kac algebra is obtained. According to [7, Theorem 4.4], the map $N \mapsto A(N)$ gives a one-to-one correspondence between the lattice of right co-ideals of \mathbb{G} and that of intermediate subfactors of $A^\alpha \subseteq A$.

Proposition 5.1. *The inverse map of ILP's Galois correspondence cited above is given by $B \mapsto N_\alpha(B)$.*

Proof. Let L, \bar{A} and $\bar{\alpha}$ be as in the proof of Theorem 3.6. Then it is easy to check that, for any intermediate subfactor B of $A^\alpha \subseteq A$ and any right co-ideal N of \mathbb{G} , one has

$$(5.1) \quad \bar{A}(N) = L \otimes A(N), \quad N_{\bar{\alpha}}(L \otimes B) = N_\alpha(B).$$

(Note that any intermediate subfactor C of $\bar{A}^{\bar{\alpha}} \subseteq \bar{A}$ has the form $C = L \otimes B$ for a unique B as above, thanks to [7, Theorem 3.9] and (the proof of) [15, Lemma 2.1].) Therefore, by considering \bar{A} instead of A itself if necessary, we may assume from the outset that A^α is infinite. Then α is dominant by [17, Theorem 2.2] or [16, Proposition 6.4]. So there exists an outer action β of $\widehat{\mathbb{G}}'$ on A^α such that $\{A, \alpha\}$ is conjugate to $\{\widehat{\mathbb{G}}' \beta \ltimes A^\alpha, \hat{\beta}\}$. Hence we assume that $A = \widehat{\mathbb{G}}' \beta \ltimes A^\alpha$ and $\alpha = \hat{\beta}$.

First, by Lemma 4.6, $N = N_\alpha(A(N))$ for any right co-ideal N of \mathbb{G} .

Let B be an intermediate subfactor of $A^\alpha \subseteq A$. Choose the unique right co-ideal N_1 such that $B = A(N_1)$ by using [7]. By the result of the preceding paragraph, we obtain $A(N_\alpha(B)) = A(N_\alpha(A(N_1))) = A(N_1) = B$. This completes the proof. \square

The following proposition is regarded as an extension of [20, Theorem 3.5], where we discussed the case of \mathbb{G} being finite-dimensional.

Proposition 5.2. *Suppose that α is a minimal action of a compact Kac algebra $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A with A^α infinite. Then $N_\alpha(\mathcal{N}(A^\alpha)') = IG(\mathbb{G})''$, i.e., the right co-ideal of \mathbb{G} corresponding to the intermediate subfactor $\mathcal{N}(A^\alpha)'$ is $IG(\mathbb{G})''$.*

Proof. Let P_1 be the basic extension of the inclusion $P := \mathcal{N}(A^\alpha)' \subseteq A$. With ρ the isomorphism introduced just after Lemma 4.1, set $Q_1 := \rho^{-1}(P_1)$. By Lemma 4.1 and Theorem 4.8, we have $\widehat{M}^{\mathcal{G}(\widehat{\mathbb{G}}^{op})} = N_{\hat{\alpha}}(Q_1)$. From this and Corollary 6.3, we obtain $\widehat{M}^{\mathcal{G}(\widehat{\mathbb{G}}^{op})} = \hat{R}(\widehat{M} \cap N_\alpha(P)')$. In other words, $\widehat{M} \cap IG(\mathbb{G}')' = \hat{R}(\widehat{M} \cap N_\alpha(P)')$. Hence, by (2.1), we have

$$\begin{aligned} N_\alpha(P) &= M \cap (\widehat{M} \cap N_\alpha(P)')' = M \cap \hat{R}(\widehat{M} \cap IG(\mathbb{G}')')' \\ &= M \cap (\widehat{M} \cap JIG(\mathbb{G}')'J)' = M \cap (\widehat{M} \cap IG(\mathbb{G}')')' = IG(\mathbb{G})''. \end{aligned}$$

Thus we are done. \square

Remark. The above proposition may be used in order to prove Theorem 4.9 in the case of compact Kac algebra actions.

6. APPENDIX

Let α be a minimal and integrable action of a locally compact quantum group $\mathbb{G} = (M, \Delta, \varphi, \psi)$ on a factor A . We fix a faithful normal semifinite weight ω on A and regard A as acting on H_ω . Denote by \tilde{J}_ω the modular conjugation of the dual weight $\tilde{\omega}$. The canonical implementation U of α associated to ω is then given by $U = \tilde{J}_\omega(\hat{J} \otimes J_\omega)$. The basic extension of $A^\alpha \subseteq A$ is denoted by $A_1 := J_\omega(A^\alpha)'J_\omega$.

Let ρ be the isomorphism $\rho: \mathbb{G}_{\alpha \ltimes} A \longrightarrow A_1$ that appeared in Section 3. Identity (4.2) can be rewritten in the form:

$$(6.1) \quad \rho(\lambda(\phi) \otimes 1) = (\phi \otimes id)(U^*) \quad (\phi \in M_*).$$

Let A_2 be the basic extension of $A \subseteq A_1$. Since $B(H_\varphi) \otimes A$ is the basic extension of $\alpha(A) \subseteq \mathbb{G}_{\alpha \ltimes} A$, the above isomorphism ρ can be extended to the isomorphism, still denoted by ρ , from $B(H_\varphi) \otimes A$ onto A_2 . Since $B(H_\varphi) \otimes A \cap (\mathbf{C} \otimes A^\alpha)' = B(H_\varphi) \otimes \mathbf{C}$, the equation

$$\Pi(T) := \rho(T \otimes 1) \quad (T \in B(H_\varphi))$$

defines a $*$ -isomorphism Π from $B(H_\varphi)$ onto $A_2 \cap (A^\alpha)'$. Since $B(H_\varphi) \otimes A \cap \alpha(A)' = M' \otimes \mathbf{C}$ and $\mathbb{G}_{\alpha \ltimes} A \cap (\mathbf{C} \otimes A^\alpha)' = \widehat{M} \otimes \mathbf{C}$, it follows from (6.1) (recall that $\widehat{M} = \{\lambda(\phi) : \phi \in M_*\}''$) that

$$(6.2) \quad A_2 \cap A' = \Pi(M'), \quad A_1 \cap (A^\alpha)' = \Pi(\widehat{M}) = \{(\phi \otimes id)(U^*) : \phi \in M_*\}''.$$

Since $\hat{R}(\lambda(\phi)) = J\lambda(\phi)^*J = \lambda(\phi \circ R)$, it follows from (6.1) that

$$(6.3) \quad \Pi(\hat{R}(z)) = J_\omega \Pi(z)^* J_\omega \quad (z \in \widehat{M}).$$

Lemma 6.1 (Proposition 4.4, [4]). *For any intermediate subfactor B of $A^\alpha \subseteq A$, we have*

$$N_\alpha(B)' \otimes \mathbf{C} = B(H_\varphi) \otimes A \cap \alpha(B)'.$$

Equivalently, $\Pi(N_\alpha(B)') = A_2 \cap B'$.

Proof. Since $B(H_\varphi) \otimes A \cap (\mathbf{C} \otimes A^\alpha)' = B(H_\varphi) \otimes \mathbf{C}$, we see that $B(H_\varphi) \otimes A \cap \alpha(B)' = \{T \otimes 1 : T \in B(H_\varphi), T \otimes 1 \in \alpha(B)'\}$. For $T \in B(H_\varphi)$, we have

$$\begin{aligned} T \otimes 1 \in \alpha(B)' &\iff (\chi \otimes \phi)((T \otimes 1)\alpha(b)) = (\chi \otimes \phi)(\alpha(b)(T \otimes 1)) \\ &\quad (\chi \in B(H_\varphi)_*, \phi \in A_*) \\ &\iff \chi(T(id \otimes \phi)(\alpha(b))) = \chi((id \otimes \phi)(\alpha(b))T) \\ &\quad (\chi \in B(H_\varphi)_*, \phi \in A_*). \end{aligned}$$

The last condition is equivalent to T being in $N_\alpha(B)'$. □

Lemma 6.2 (Corollaire 4.5, [4]). *Let B be an intermediate subfactor of $A^\alpha \subseteq A$.*

- (1) *We have $\widehat{M} \cap N_\alpha(B)' \otimes \mathbf{C} = \mathbb{G}_{\alpha \ltimes} A \cap \alpha(B)'$, i.e., $\Pi(\widehat{M} \cap N_\alpha(B)') = A_1 \cap B'$.*
- (2) *We have $N_\alpha(B)' \otimes \mathbf{C} = (M' \otimes \mathbf{C}) \vee (\mathbb{G}_{\alpha \ltimes} A \cap \alpha(B)'),$ i.e., $A_2 \cap B' = (A_2 \cap A') \vee (A_1 \cap B').$*
- (3) *Let $B_1 := J_\omega B' J_\omega$ be the basic extension of $B \subseteq A$, and set $\tilde{B}_1 := \rho^{-1}(B_1)$. Then $\Pi(\hat{R}(\widehat{M} \cap N_\alpha(B)')) = B_1 \cap (A^\alpha)'$. In particular, $\tilde{B}_1 \cap (\mathbf{C} \otimes A^\alpha)' = \hat{R}(\widehat{M} \cap N_\alpha(B)') \otimes \mathbf{C}.$*

Proof. (1) This follows from Lemma 6.1 and the identity $\mathbb{G}_{\alpha} \ltimes A \cap (\mathbf{C} \otimes A^{\alpha})' = \widehat{M} \otimes \mathbf{C}$.

(2) This is due to the fact that $N_{\alpha}(B) = M \cap (\widehat{M} \cap N_{\alpha}(B)')'$.

(3) This follows from Part (1) and (6.3). \square

Theorem 6.3. *Let B, B_1, \tilde{B}_1 be as in Lemma 6.2. If \mathbb{G} is a compact Kac algebra, then we have $B_1 = A \vee (B_1 \cap (A^{\alpha})')$ and $\tilde{B}_1 = \alpha(A) \vee \hat{R}(\widehat{M} \cap N_{\alpha}(B)') \otimes \mathbf{C}$. Moreover, $N_{\hat{\alpha}}(\tilde{B}_1) = \hat{R}(\widehat{M} \cap N_{\alpha}(B)')$.*

Proof. For ω , choose a faithful normal state ω_0 on A^{α} and put $\omega := \omega_0 \circ T_{\alpha}$. By [7, Theorem 3.9], there exists a (unique) conditional expectation E_B from A onto B . Let e_B be the Jones projection of B : $e_B \Lambda_{\omega}(a) = \Lambda_{\omega}(E_B(a))$. So $B_1 = (A \cup \{e_B\})''$. Since $e_B \in B_1 \cap (A^{\alpha})'$, we find that $B_1 = A \vee (B_1 \cap (A^{\alpha})')$. From Lemma 6.2 (3), it follows that $\tilde{B}_1 = \alpha(A) \vee \hat{R}(\widehat{M} \cap N_{\alpha}(B)') \otimes \mathbf{C}$. From Lemma 4.6, we find that $N_{\hat{\alpha}}(\tilde{B}_1) = \hat{R}(\widehat{M} \cap N_{\alpha}(B)')$. \square

REFERENCES

- [1] H. Araki, R. Haag, D. Kastler and M. Takesaki, Extension of KMS states and chemical potential, *Commun. Math. Phys.*, **53** (1977), 97–134. MR **56**:2042
- [2] J. De Cannière, On the intrinsic group of a Kac algebra, *Proc. London Math. Soc.*, **40** (1980), 1–20. MR **83d**:46079
- [3] J. Dixmier and O. Maréchal, Vecteurs totalisateurs d'une algèbre de von Neumann, *Commun. Math. Phys.*, **22** (1971), 44–50. MR **45**:5767
- [4] M. Enock, Sous-facteurs intermédiaires et groupes quantiques mesurés, *J. Operator Theory*, **42** (1999), 305–330. MR **2000i**:46057
- [5] M. Enock and J.-M. Schwartz, Systèmes dynamiques généralisés et correspondances, *J. Operator Theory*, **11** (1984), 273–303. MR **86g**:46097
- [6] M. Enock and J.-M. Schwartz, *Kac Algebras and Duality of Locally Compact Groups*, Springer-Verlag, Berlin (1992). MR **94e**:46001
- [7] M. Izumi, R. Longo and S. Popa, A Galois correspondence for compact groups of automorphisms of von Neumann algebras with a generalization to Kac algebras, *J. Funct. Anal.*, **155** (1998), 25–63. MR **2000c**:46117
- [8] H. Kosaki, Extension of Jones' theory on index to arbitrary factors, *J. Funct. Anal.*, **66** (1986), 123–140. MR **87g**:46093
- [9] J. Kustermans and S. Vaes, Locally compact quantum groups, *Ann. Sci. École Norm. Sup.*, 4^e série, t. **33** (2000), 837–934. MR **2002f**:46108
- [10] J. Kustermans and S. Vaes, Locally compact quantum groups in the von Neumann algebra setting, To appear in *Math. Scand.*
- [11] M. Pimsner and S. Popa, Entropy and index for subfactors, *Ann. Sci. École Norm. Sup.*, 4^e série, t. **19** (1986), 57–106. MR **87m**:46120
- [12] J.-M. Schwartz, Sur la structure des algèbres de Kac, I, *J. Funct. Anal.*, **34** (1979), 370–406. MR **83a**:46072a
- [13] Ș. Strătilă, *Modular theory in operator algebras*, Abacus Press, Tunbridge Wells, 1981. MR **85g**:46072
- [14] M. Takesaki, The structure of a von Neumann algebra with a homogeneous periodic state, *Acta Math.*, **131** (1973), 79–121. MR **55**:11067
- [15] T. Teruya and Y. Watatani, Lattices of intermediate subfactors for type III factors, *Arch. Math.*, **68** (1997), 454–463. MR **98d**:46071
- [16] S. Vaes, The unitary implementation of a locally compact quantum group action, *J. Funct. Anal.*, **180** (2001), 426–480. MR **2002a**:46100
- [17] T. Yamanouchi, On dominance of minimal actions of compact Kac algebras and certain automorphisms in $\text{Aut}(A/A^{\alpha})$, *Math. Scand.*, **84** (1999), 297–319. MR **2000g**:46101
- [18] T. Yamanouchi, The Connes spectrum for actions of compact Kac algebras and factoriality of their crossed products, *Hokkaido Math. J.*, **28** (1999), 409–434. MR **2000g**:46100

- [19] T. Yamanouchi, Uniqueness of Haar measures for a quasi Woronowicz algebra, *Hokkaido Math. J.*, **30** (2001), 105–112. MR **2002e:46075**
- [20] T. Yamanouchi, Description of the automorphism group $\text{Aut}(\mathcal{A}/\mathcal{A}^\alpha)$ for a minimal action of a compact Kac algebra and its application, *J. Funct. Anal.*, **194** (2002), 1–16.

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, HOKKAIDO UNIVERSITY, SAPPORO 060-0810 JAPAN

E-mail address: yamanouc@math.sci.hokudai.ac.jp

COMPOSITION OPERATORS ACTING ON HOLOMORPHIC SOBOLEV SPACES

BOO RIM CHOE, HYUNGWOON KOO, AND WAYNE SMITH

ABSTRACT. We study the action of composition operators on Sobolev spaces of analytic functions having fractional derivatives in some weighted Bergman space or Hardy space on the unit disk. Criteria for when such operators are bounded or compact are given. In particular, we find the precise range of orders of fractional derivatives for which all composition operators are bounded on such spaces. Sharp results about boundedness and compactness of a composition operator are also given when the inducing map is polygonal.

1. INTRODUCTION AND STATEMENT OF RESULTS

Let \mathbf{D} be the unit disk in the complex plane. We shall write $H(\mathbf{D})$ for the class of all holomorphic functions on \mathbf{D} . Let $s \geq 0$ be a real number. Following [BB], we define the fractional derivative for $f \in H(\mathbf{D})$ of order s by

$$\mathcal{R}^s f(z) = \sum_{n=0}^{\infty} (1+n)^s a_n z^n, \quad z \in \mathbf{D}$$

where $\sum_{n=0}^{\infty} a_n z^n$ is the Taylor series of f .

In this paper, we are going to investigate composition operators acting on holomorphic Sobolev spaces defined in terms of fractional derivatives. To introduce those holomorphic Sobolev spaces, let us first recall some well-known function spaces. For $0 < p < \infty$ and $\alpha > -1$, the weighted Bergman space A_{α}^p is the space of all $f \in H(\mathbf{D})$ for which

$$\|f\|_{A_{\alpha}^p}^p = \int_{\mathbf{D}} |f(z)|^p (1-|z|^2)^{\alpha} dA(z) < \infty,$$

where dA is area measure on \mathbf{D} . Also, the Hardy space H^p is the space of all $g \in H(\mathbf{D})$ for which

$$\|g\|_{H^p}^p = \sup_{0 < r < 1} \int_0^{2\pi} |g(re^{i\theta})|^p \frac{d\theta}{2\pi} < \infty.$$

We will often use the following notation to allow unified statements:

$$A_{-1}^p = H^p.$$

This notation is justified by the weak-star convergence of $(\alpha+1)(1-|z|^2)^{\alpha} dA(z)/\pi$ to $d\theta/2\pi$ as $\alpha \rightarrow -1$.

Received by the editors April 4, 2002 and, in revised form, August 6, 2002.

2000 *Mathematics Subject Classification.* Primary 47B33; Secondary 30D55, 46E15.

Key words and phrases. Composition operator, fractional derivative, Bergman space.

The second author's research was partially supported by KRF2001-041-D00012.

Now, for $p > 0$, $s \geq 0$ and $\alpha \geq -1$, the holomorphic Sobolev space $A_{\alpha,s}^p$ is defined to be the space of all $f \in H(\mathbf{D})$ for which $\mathcal{R}^s f \in A_\alpha^p$. We will often write $H_s^p = A_{-1,s}^p$. We define the norm of $f \in A_{\alpha,s}^p$ by

$$\|f\|_{A_{\alpha,s}^p} = \|\mathcal{R}^s f\|_{A_\alpha^p}.$$

Of course, we are abusing the term “norm”, since $\|\cdot\|_{A_{\alpha,s}^p}$ does not satisfy the triangle inequality for $0 < p < 1$, but in this case $(f, g) \mapsto \|f - g\|_{A_{\alpha,s}^p}^p$ defines a translation-invariant metric on A_α^p which turns A_α^p into a complete topological vector space.

A function $\varphi \in H(\mathbf{D})$ that satisfies $\varphi(\mathbf{D}) \subset \mathbf{D}$ induces the composition operator C_φ , defined on $H(\mathbf{D})$ by

$$C_\varphi f = f \circ \varphi.$$

Throughout this paper the symbol φ will always represent a holomorphic self-map of \mathbf{D} . In this paper we study the action of composition operators on holomorphic Sobolev spaces. This setting allows a unified treatment of composition operators on Hardy spaces ($H^p = A_{-1,0}^p$), weighted Bergman spaces ($A_\alpha^p = A_{\alpha,0}^p$, $\alpha > -1$), and Dirichlet-type spaces ($A_{\alpha,1}^p$), where extensive research has already been done. The book [CM] is a good introduction to this work. The main results in this paper may be viewed as summarizing well-known boundedness and compactness results for composition operators on these spaces, and then extending them to the Sobolev setting.

It is a well-known consequence of Littlewood’s Subordination Principle that every composition operator is bounded on A_α^p for every $p > 0$ and $\alpha \geq -1$; see [MS]. It is natural to ask how this extends to the spaces $A_{\alpha,s}^p$ when $s > 0$. For $p > 0$, $\alpha_j > -1$, $s_j \geq 0$ ($j = 1, 2$) with $\alpha_1 - \alpha_2 = p(s_1 - s_2)$, we have the following equivalence (see Theorem 5.12 in [BB]):

$$(1.1) \quad A_{\alpha_1,s_1}^p \approx A_{\alpha_2,s_2}^p.$$

That is, these spaces are isomorphic and have equivalent norms. In particular, when $s < \frac{\alpha+1}{p}$ we have $A_{\alpha,s}^p \approx A_{\alpha-sp}^p$. Thus it follows that every composition operator is bounded on $A_{\alpha,s}^p$ when $s < \frac{\alpha+1}{p}$. The general situation is described in the following theorem. Just the statement of this and our other main results are given in this section. The proofs will come later.

Theorem 1.1. *Let $p > 0$, $s \geq 0$ and $\alpha \geq -1$.*

- (a) *If $s < \frac{\alpha+1}{p}$, then every composition operator is bounded on $A_{\alpha,s}^p$.*
- (b) *If $s = \frac{\alpha+1}{p}$ and*
 - (i) *$p \geq 2$ or $\alpha = -1$, then every composition operator is bounded on $A_{\alpha,s}^p$.*
 - (ii) *$p < 2$ and $\alpha > -1$, then some composition operators are not bounded on $A_{\alpha,s}^p$.*
- (c) *If $s > \frac{\alpha+1}{p}$, then some composition operators are not bounded on $A_{\alpha,s}^p$.*

The case $\alpha = -1$ in part (b) corresponds to $s = 0$, and as previously mentioned every composition operator is bounded on $H^p = A_{-1,0}^p$. The case $\alpha = -1$ in part (c) shows that this does not extend to H_s^p for a range of positive s , as was the case for the Bergman-Sobolev spaces.

The bounds on s in Theorem 1.1 can be extended when the inducing map of the composition operator is univalent or, more generally, of bounded valence.

Theorem 1.2. *Let $p > 0$, $s \geq 0$ and $\alpha > -1$. Assume that φ is of bounded valence.*

- (a) *If $p \geq 2$ and $s \leq \frac{\alpha+2}{p}$, then C_φ is bounded on $A_{\alpha,s}^p$.*
- (b) *If $p < 2$ and $s < \frac{\alpha+1}{p} + \frac{1}{2}$, then C_φ is bounded on $A_{\alpha,s}^p$.*

The upper bound $s \leq \frac{\alpha+2}{p}$ in part (a), for $p \geq 2$, is sharp; an example will be given in §6. We do not know whether the upper bound $s < \frac{\alpha+1}{p} + \frac{1}{2}$ in part (b) is sharp, but another example will be given that shows that the upper bound cannot be extended to the bound $\frac{\alpha+2}{p}$ from part (a).

The equivalence (1.1) does not extend to the limiting case $\alpha_2 = -1$. However, for $\alpha_1 \geq -1$ with $\alpha_1 + 1 = p(s_1 - s_2)$, we have the following Littlewood-Paley-type inclusion relations:

$$(1.2) \quad p \leq 2 \implies A_{\alpha_1, s_1}^p \subset A_{-1, s_2}^p,$$

$$(1.3) \quad p \geq 2 \implies A_{-1, s_2}^p \subset A_{\alpha_1, s_1}^p.$$

Inclusion relations for different values of p are also known. For $0 < p_1 < p_2$, $\alpha_j \geq -1$, $s_j \geq 0$ ($j = 1, 2$) with $\frac{\alpha_1+2}{p_1} - \frac{\alpha_2+2}{p_2} = s_1 - s_2$, we have

$$(1.4) \quad A_{\alpha_1, s_1}^{p_1} \subset A_{\alpha_2, s_2}^{p_2}.$$

Let $p > 0$, $\alpha \geq -1$, and $s \geq 0$. Note that we have $A_{\alpha,s}^p \subset A_\alpha^{p(\alpha+2)/(\alpha+2-ps)}$ for $ps < \alpha + 2$, as a special case of the above inclusion ($s_2 = 0$, $\alpha_1 = \alpha_2$). In case $ps \geq \alpha + 2$, inclusion relations with other types of function spaces are known as follows:

$$(1.5) \quad 0 < ps - (\alpha + 2) < p \implies A_{\alpha,s}^p \subset \Lambda_{s-(\alpha+2)/p},$$

$$(1.6) \quad ps = \alpha + 2 \implies A_{\alpha,s}^p \subset \text{VMOA}.$$

Here, Λ_ε denotes the holomorphic Lipschitz space of order ε , $0 < \varepsilon < 1$, and VMOA denotes the space of holomorphic functions of vanishing mean oscillation. The definitions and more information on these spaces can be found in [CM] for Λ_ε and [G] for VMOA. For details of all the inclusions mentioned above, see Theorem 5.12, Theorem 5.13, and Theorem 5.14 in [BB].

The boundedness (compactness) of a composition operator on a smaller space often implies the boundedness (compactness) of the operator on larger spaces. This general philosophy and the inclusion relations mentioned above lead to natural conjectures. The methods developed below in §2 to address these conjectures require some restriction on the parameters. In particular, the case (1.2) is left open since our methods do not apply when the target space is a Hardy-Sobolev space.

Theorem 1.3. *Let $X \subset Y$ be any of the inclusion relations in (1.3) – (1.5), and assume for inclusion (1.3) that $s_2 < 1$, and for inclusion (1.4) that $\alpha_2 > -1$ and $s_2 < 1 + (1 + \alpha_2)/p_2$.*

- (a) *If $C_\varphi : X \rightarrow X$ is bounded, then $C_\varphi : Y \rightarrow Y$ is bounded.*
- (b) *If $C_\varphi : X \rightarrow X$ is compact, then $C_\varphi : Y \rightarrow Y$ is compact.*

Inclusion (1.6) was left out of the preceding theorem, but we have the following partial result in that case.

Theorem 1.4. *Let $p > 0$, $\alpha \geq -1$, $s \geq 0$ and assume $ps = \alpha + 2$.*

- (a) *If $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is bounded, then $C_\varphi : \text{VMOA} \rightarrow \text{VMOA}$ is bounded.*

- (b) If φ is univalent and $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is compact, then $C_\varphi : \text{VMOA} \rightarrow \text{VMOA}$ is compact.

We also mention the elementary inclusion relations that, for all $p > 0$, $s \geq 0$, $\alpha \geq -1$, and $\varepsilon \geq 0$,

$$A_{\alpha,s+\varepsilon}^p \subset A_{\alpha,s}^p \subset A_{\alpha+\varepsilon,s}^p.$$

In §3 we will give a result analogous to Theorem 1.3 for these inclusions, but with some restrictions on the parameters.

As a first application of Theorem 1.3, notice that it can be used to prove the case $\alpha > -1$, $p \geq 2$, and $s = \frac{\alpha+1}{p}$ in Theorem 1.1. Then $H^p \subset A_{\alpha,s}^p$ by (1.3), and so every composition operator is bounded on $A_{\alpha,s}^p$ by Theorem 1.3. In the other direction, once criteria for C_φ to be bounded or compact on the larger spaces are known, Theorem 1.3 can be used to provide necessary conditions for boundedness or compactness of C_φ on the smaller spaces. For example, by taking Λ_ε as the larger space, we have the following consequence, which has been known for $p \geq 2$ (Theorem 4.13 in [CM]), while it has been known to be false for $p = 1$ (p. 193 in [CM]). So, the gap $1 < p < 2$ is now filled in. A more general version is proved as Theorem 3.3 below.

Theorem 1.5. *Let $p > 1$ and suppose $C_\varphi : H_1^p \rightarrow H_1^p$ is bounded. Then the angular derivative of φ exists at all points $\zeta \in \partial\mathbf{D}$ where the radial limit $\varphi(\zeta)$ of φ exists and satisfies $|\varphi(\zeta)| = 1$.*

A basic problem in the study of composition operators is to relate function-theoretic properties of φ to operator-theoretic properties of the restriction of C_φ to various spaces, as in Theorem 1.5. When $ps < \alpha + 1$, we have $A_{\alpha,s}^p \approx A_{\alpha-sp}^p$ by (1.1), and criteria for $C_\varphi : A_{\alpha_1}^{p_1} \rightarrow A_{\alpha_2}^{p_2}$ to be bounded or compact are known. The characterization is that a generalized Nevanlinna counting function for φ satisfies a growth condition if $p_2 \geq p_1$, or an integrability condition if $p_2 < p_1$; see [Sm1] and [SY]. The results in [Sm1] and [SY] do not apply when $ps \geq \alpha + 1$ in either the domain or the target space. In that case, criteria in the form of Carleson measure conditions for a measure defined using a modified counting function can be obtained as in Theorem 2.6 below, with some restrictions on the parameters α_j , p_j , and s_j . This Carleson-type criteria in Theorem 2.6 will be used to prove Theorems 1.2 and 1.3. We also mention that for the special case $p = 2$ other techniques are available, since the norm of a function in $A_{\alpha,s}^2$ may be given in terms of its power series coefficients. These spaces are examples of what are called weighted Hardy spaces in [CM], which is a good reference for composition operators acting on these spaces.

Characterizing when a composition operator is bounded on H_s^p , $s > 0$, seems much harder. The difficulty is that (1.1) does not provide an isomorphism with a space of functions defined with full derivatives, and the methods used to prove Theorem 2.6 do not apply. We have from Theorem 1.1 that, for any $p > 0$ and $s > 0$, there exists a function φ such that C_φ is not bounded on H_s^p . A positive result is that C_φ is compact on certain H_s^p whenever φ is of bounded valence and $\varphi(\mathbf{D})$ is contained in a polygonal region contained in \mathbf{D} . This is the special case $p_1 = p_2$ of the following result. For a polygon P inscribed in the unit circle, let $\theta(P)$ denote $1/\pi$ times the measure of the largest vertex angle of P .

Theorem 1.6. *Let $p_2 \geq p_1 > 0$ and assume $0 \leq s < \min\{\frac{1}{p_2}, \frac{1}{2}\}$. Let φ be a holomorphic function of bounded valence taking \mathbf{D} into a polygon P inscribed in the unit circle. If $\theta(P) < \frac{p_1(1-sp_2)}{p_2(1-sp_1)}$, then $C_\varphi : H_s^{p_1} \rightarrow H_s^{p_2}$ is compact.*

When $s = 0$ and $p_1 = p_2$, this has long been known; see [ST]. When $s = 0$ and $p_2 \geq p_1$, this result is basically contained in [Sm1]. These results (when $s = 0$) do not require the hypothesis of bounded valence. We will prove a more general result in Theorem 5.5.

In the next section we develop the change of variable methods that we use to study composition operators, which we then use to give Carleson measure-type criteria for these operators to be bounded or compact. These criteria are then used in §3 to prove Theorem 1.3. Next, in §4, the proofs of Theorem 1.1 and Theorem 1.2 are given. Simple geometric criteria are then developed in §5 for boundedness and compactness of a composition operator between holomorphic Sobolev spaces when the inducing map is polygonal. The paper concludes, in §6, with several examples which demonstrate that our theorems are sharp.

2. BACKGROUND: CARLESON-TYPE CRITERIA

Our approach to studying composition operators on the spaces $A_{\alpha,s}^p$ involves a change of variable from z to $w = \varphi(z)$. The equivalence (1.1) allows us to assume that s is an integer, and then standard non-univalent change of variable methods can be applied. This gets quite complicated when s is an integer greater than 1. Thus, for simplicity and clarity of presentation, we confine our attention to the case $s = 1$. This enables us to cover parameters p , α and s with $\alpha + (1-s)p > -1$ by using the equivalence $A_{\alpha,s}^p \approx A_{\alpha+(1-s)p,1}^p$ from (1.1). The change of variable method for $s = 1$ is summarized as follows.

For a holomorphic map $\varphi : \mathbf{D} \rightarrow \mathbf{D}$ and $w \in \mathbf{D}$, define the modified counting function $N_{p,\alpha}(\varphi, w)$ corresponding to the measure $(1 - |z|^2)^\alpha dA(z)$ by

$$N_{p,\alpha}(\varphi, w) = \sum |\varphi'(z)|^{p-2} (1 - |z|^2)^\alpha$$

where the sum is over the set $\{z : \varphi(z) = w\}$. As usual, the zeros of $\varphi - w$ are repeated according to their multiplicity. The change of variable formula we need uses the measure

$$d\mu_{p,\alpha}^\varphi(w) = N_{p,\alpha}(\varphi, w) dA(w).$$

Then, by the area formula (see Theorem 2.32 in [CM]), we have the following change of variable formula.

Proposition 2.1. *Let $p > 0$ and $\alpha > -1$. Then, we have*

$$\int_{\mathbf{D}} |(f \circ \varphi)'(z)|^p (1 - |z|^2)^\alpha dA(z) = \int_{\varphi(\mathbf{D})} |f'(w)|^p d\mu_{p,\alpha}^\varphi(w)$$

for functions $f \in H(\mathbf{D})$.

Note that Proposition 2.1 cannot be directly applied to the case $s = 1$, because $\mathcal{R}f(z) = f(z) + zf'(z)$ by our definition. This difficulty is overcome by the following proposition. We will often write $X \lesssim Y$ if $X \leq CY$ for some positive constant C dependent only on allowed parameters, and $X \approx Y$ if $X \lesssim Y \lesssim X$.

Proposition 2.2. *Let $p > 0$, $\alpha \geq -1$ and $a \in \mathbf{D}$. Then, for every positive integer n , we have*

$$\|f\|_{A_{\alpha,n}^p} \approx \sum_{k=0}^{n-1} |f^{(k)}(a)| + \|f^{(n)}\|_{A_{\alpha}^p}$$

for $f \in H(\mathbf{D})$.

Proof. We prove the proposition for $a = 0$. The proof for general a is similar. The equivalence $\|f\|_{A_{\alpha,n}^p} \approx \sum_{k=0}^n \|f^{(k)}\|_{A_{\alpha}^p}$ is proved in Theorem 5.3 of [BB]. Thus, $\|f\|_{A_{\alpha,n}^p} \gtrsim \sum_{k=0}^{n-1} |f^{(k)}(0)| + \|f^{(n)}\|_{A_{\alpha}^p}$ is clear by subharmonicity.

Now, we prove the other direction of the inequalities. Since $H(\overline{\mathbf{D}})$ is dense in all holomorphic Sobolev spaces by Lemma 5.2 of [BB], it is sufficient to show that

$$(2.1) \quad \|f\|_{A_{\alpha}^p} \lesssim |f(0)| + \|f'\|_{A_{\alpha}^p}, \quad f \in H(\overline{\mathbf{D}}).$$

First, assume either $\alpha > -1$ or $0 < p \leq 1$. Let $f \in H(\overline{\mathbf{D}})$. For each $\beta > -1$, we have by Theorem 1.9 of [BB],

$$f(z) = \frac{1}{\pi} \int_{\mathbf{D}} \mathcal{R}f(w) G_{\beta}(z\bar{w})(1 - |w|^2)^{\beta} dA(w)$$

where

$$G_{\beta}(z) = \frac{1}{z} \left\{ \frac{1}{(1-z)^{1+\beta}} - 1 \right\}.$$

Therefore, choosing $\beta > -1$ sufficiently large, we have by Lemma 4.1 of [BB] ($\alpha > -1$ or $0 < p \leq 1$ is used here),

$$(2.2) \quad \|f\|_{A_{\alpha}^p} \lesssim \|\mathcal{R}f\|_{A_{\alpha+p}^p} \approx \|f\|_{A_{\alpha+p}^p} + \|f'\|_{A_{\alpha+p}^p}.$$

It is easy to see that, given $\varepsilon > 0$, there exist a constant $C > 0$ and a compact subset $K = \{z \in \mathbf{D} : |z| \leq r < 1\}$ of \mathbf{D} such that

$$\|f\|_{A_{\alpha+p}^p} \leq \varepsilon \|f\|_{A_{\alpha}^p} + C \sup_{z \in K} |f(z)|.$$

Taking $\varepsilon > 0$ sufficiently small, we have by (2.2),

$$\begin{aligned} \|f\|_{A_{\alpha}^p} &\lesssim \|f'\|_{A_{\alpha+p}^p} + \sup_{z \in K} |f(z)| \\ &\lesssim |f(0)| + \|f'\|_{A_{\alpha}^p} + \sup_{z \in K} |f(z) - f(0)| \\ &\lesssim |f(0)| + \|f'\|_{A_{\alpha}^p} + \sup_{z \in K} |f'(z)|. \end{aligned}$$

Since $\sup_{z \in K} |f'(z)| \lesssim \|f'\|_{A_{\alpha}^p}$ by the subharmonicity of $|f'|^p$, we obtain (2.1) as desired.

Now, consider the case $\alpha = -1$ and $p > 1$. Note that

$$|f(e^{i\theta}) - f(0)| \leq \int_0^1 |f'(te^{i\theta})| dt.$$

Therefore, by Minkowski's inequality, we have

$$\|f - f(0)\|_{H^p} \leq \int_0^1 \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f'(te^{i\theta})|^p d\theta \right\}^{1/p} dt \leq \|f'\|_{H^p},$$

which implies (2.1). The proof is complete. \square

Having seen Proposition 2.1 and Proposition 2.2, it is now clear that the behavior of C_φ , when the target space is $A_{\alpha,1}^p$, depends on that of the measure $\mu_{p,\alpha}^\varphi$. For boundedness and compactness of C_φ , the criteria for $\mu_{p,\alpha}^\varphi$ turn out to be Carleson-type conditions in certain cases. To prove it, we need a couple of lemmas.

Lemma 2.3. *A bounded subset of any of the spaces $A_{\alpha,s}^p$, where $p > 0$, $s \geq 0$, and $\alpha \geq -1$, is a normal family.*

Proof. First assume $\alpha > -1$. Using (1.1), a bounded set X in $A_{\alpha,s}^p$ is also bounded in some $A_{\beta,n}^p$, where n is a nonnegative integer. Recall that there is a constant C such that

$$|g(w)| \leq C \|g\|_{A_\beta^p} (1 - |w|)^{-(\beta+2)/p}$$

for all $g \in A_\beta^p$ (see, for example, Theorem 7.2.5 in [R1]). By Proposition 2.2, this shows that the functions in X are uniformly bounded on compact subsets of \mathbf{D} . Hence X is a normal family. The proof of the result for $\alpha = -1$ is similar, since $A_{-1,s}^p \subset A_{0,s}^{2p}$ by (1.4). The proof is complete. \square

In the next lemma we will need the estimate that if $\alpha > -1$ and $\beta > 0$, then

$$(2.3) \quad \int_{\mathbf{D}} \frac{(1 - |z|^2)^\alpha}{|1 - \bar{a}z|^{2+\alpha+\beta}} dA(z) \approx \frac{1}{(1 - |a|^2)^\beta} \quad (|a| \rightarrow 1-).$$

A reference is Theorem 1.7 of [HKZ].

Lemma 2.4. *Let $p > 0$, $\alpha \geq -1$ and $s \geq 0$. Let $N > \frac{\alpha+2}{p} - s$. Put $g_a(z) = (1 - z\bar{a})^{-N}$ for $a, z \in \mathbf{D}$. Then, we have*

$$\|g_a\|_{A_{\alpha,s}^p} \approx (1 - |a|)^{-N-s+\frac{\alpha+2}{p}}, \quad a \in \mathbf{D}$$

where the constants in this estimate depend on N , s , α , and p , but are independent of a .

Proof. First, assume $\alpha > -1$. Let k be the smallest integer satisfying $k \geq s$. Then, we have $A_{\alpha,s}^p \approx A_{\alpha+(k-s)p,k}^p$ by (1.1). Thus, by Proposition 2.2, we have

$$\|g_a\|_{A_{\alpha,s}^p} \approx 1 + \sum_{j=1}^{k-1} c_{N,j} |a|^j + c_{N,k} |a|^k \left\{ \int_{\mathbf{D}} \frac{(1 - |z|^2)^{\alpha+(k-s)p}}{|1 - z\bar{a}|^{p(N+k)}} dA(z) \right\}^{1/p}$$

where $c_{N,j} = N(N+1) \dots (N+j-1)$. Thus, by (2.3), we have

$$\|g_a\|_{A_{\alpha,s}^p} \approx 1 + \sum_{j=1}^{k-1} c_{N,j} |a|^j + c_{N,k} C_1 |a|^k (1 - |a|)^{-N-s+\frac{\alpha+2}{p}},$$

where $C_1 = C_1(N, s, \alpha, p)$. The desired estimate follows. Next, assume $\alpha = -1$. Note that $A_{0,3/p+s}^{p/2} \subset A_{-1,s}^p \subset A_{0,s}^{2p}$ by (1.4) and thus

$$\|g_a\|_{A_{0,s}^{2p}} \lesssim \|g_a\|_{A_{-1,s}^p} \lesssim \|g_a\|_{A_{0,3/p+s}^{p/2}}.$$

On the other hand, we have

$$\|g_a\|_{A_{0,s}^{2p}} \approx (1 - |a|)^{-N-s+\frac{1}{p}} \approx \|g_a\|_{A_{0,3/p+s}^{p/2}}$$

by what we have just proved for the case $\alpha > -1$. This completes the proof. \square

For any arc $I \subset \partial \mathbf{D}$ define the Carleson square over I to be

$$SI = \{re^{i\theta} : 1 - |I| \leq r < 1, e^{i\theta} \in I\},$$

where $|I|$ is $1/(2\pi)$ times the Euclidean length of I . Also, let ∂ denote the complex differential operator, i.e., $\partial f = f'$ for $f \in H(\mathbf{D})$.

The next lemma asserts that certain operators are compact. We review the definition, since when $p < 1$ the spaces involved are not Banach spaces. Suppose X and Y are complete topological vectors spaces whose topologies are induced by metrics. A continuous linear operator $T : X \rightarrow Y$ is said to be compact if the image of every bounded set in X is relatively compact in Y . Due to the metric topology of Y , T will be compact if and only if the image of every bounded sequence in X has a subsequence that converges in Y . Also, linearity of T allows us to only consider sequences in the unit ball of X .

In the following lemma, part (a) is well known; see Theorems 2.2 and 3.1 in [L1]. Part (b) is certainly known to experts. For example, the case $k = 0$, $p = q$, and $\alpha > -1$ occurs as Theorem 4.3 in [MS]. A proof is included here since we do not know a reference. In our application, we will take $k \leq 1$.

Lemma 2.5. *Assume that one of the following three conditions holds:*

- (i) $\alpha > -1, 0 < p \leq q$; (ii) $\alpha = -1, p = q \geq 2$; (iii) $\alpha = -1, 0 < p < q$.

Let k be a nonnegative integer and μ be a positive finite Borel measure on \mathbf{D} .

- (a) $\partial^k : A_\alpha^p \rightarrow L^q(d\mu)$ is bounded if and only if

$$\mu(SI) = O\left(|I|^{kq+q(\alpha+2)/p}\right), \quad I \subset \partial \mathbf{D}.$$

- (b) $\partial^k : A_\alpha^p \rightarrow L^q(d\mu)$ is compact if and only if

$$(2.4) \quad \mu(SI) = o\left(|I|^{kq+q(\alpha+2)/p}\right), \quad |I| \rightarrow 0.$$

Moreover, the norm of the map in (a) satisfies the inequality $\|\partial^k\|^q \leq C\|\mu\|$, where $\|\mu\|$ is the supremum of the quantity $\mu(SI)/|I|^{kq+q(\alpha+2)/p}$ over $I \subset \partial \mathbf{D}$.

Proof. We provide a proof of (b). We first prove the sufficiency. So, assume (2.4) and let $\{f_n\}$ be a bounded sequence in A_α^p , say of norm at most $1/2$. We must show that $\{f_n\}$ contains a subsequence whose k -th derivatives converge in $L^q(d\mu)$. Recall that we have observed that a bounded set in A_α^p is a normal family, and so by subtracting the limit function and re-indexing an appropriate subsequence, we may assume that $\|f_n\|_{A_\alpha^p} \leq 1$ and that $\{f_n\}$ and hence $\{f_n^{(k)}\}$ converges to 0 uniformly on compact subsets of \mathbf{D} . We need to show that $\{f_n^{(k)}\}$ converges to 0 in $L^q(d\mu)$. Let $\varepsilon > 0$ and write

$$\|f_n^{(k)}\|_{L^q(d\mu)}^q = \int_{r\mathbf{D}} |f_n^{(k)}|^q d\mu + \int_{\mathbf{D} \setminus r\mathbf{D}} |f_n^{(k)}|^q d\mu, \quad 0 < r < 1.$$

The first term is easily handled. For any fixed $r \in (0, 1)$, the uniform convergence of $\{f_n^{(k)}\}$ to 0 on $r\mathbf{D}$ allows us to find $N(r)$ such that

$$\int_{r\mathbf{D}} |f_n^{(k)}|^q d\mu < \varepsilon, \quad n \geq N(r).$$

Turning to the second term, by hypothesis we can choose $r = r_\varepsilon \in (0, 1)$ so that the measure $d\nu(w) = \chi_{\mathbf{D} \setminus r\mathbf{D}}(w) d\mu(w)$ satisfies $\nu(SI) \leq \varepsilon |I|^\beta$, whenever $|I| \leq 1 - r$,

where $\beta = kq + q(2 + \alpha)/p$. For $|I| > 1 - r$, we subdivide I into m arcs of length at most $1 - r$, where $m \leq |I|/(1 - r) + 1 \leq 2|I|/(1 - r)$, and observe that $SI \cap (\mathbf{D} \setminus r\mathbf{D})$ is contained in the Carleson squares associated with the smaller arcs. Thus, the previous estimate shows that

$$\nu(SI) \leq \frac{2|I|}{1-r} \varepsilon (1-r)^\beta \leq 2\varepsilon |I|^\beta$$

in this case as well. Note that we used $\beta \geq 1$, which is a consequence of the hypotheses, for the last inequality. Thus, we see from (a) that there is a constant C_1 such that

$$\sup_n \int_{\mathbf{D}} |f_n^{(k)}|^q d\nu \leq C_1 \varepsilon.$$

Combined with the previous estimate, this shows that $\|f_n^{(k)}\|_{L^q(d\mu)} \rightarrow 0$ as required.

Now, we prove the necessity. Suppose (2.4) is false. Then there exist a constant $C_2 > 0$ and a sequence of arcs $I_n \subset \partial\mathbf{D}$ such that $|I_n| \rightarrow 0$ and

$$(2.5) \quad \mu(SI_n) \geq C_2 |I_n|^{kq+q(\alpha+2)/p}.$$

Let $\delta_n = |I_n|$ and $\zeta_n \in \partial\mathbf{D}$ be the center of I_n for each n . Fix a large integer $N > (\alpha + 2)/p$. Let $g_n(z) = (1 - (1 - \delta_n)z\bar{\zeta}_n)^{-N}$ and put $f_n = g_n \|g_n\|_{A_\alpha^p}^{-1}$. Note that $\|g_n\|_{A_\alpha^p}^p \approx \delta_n^{-Np+\alpha+2}$ by Lemma 2.4. Thus, $\{f_n\}$ converges uniformly to 0 on compact subsets of \mathbf{D} . Now, using the compactness of $\partial^k : A_\alpha^p \rightarrow L^q(d\mu)$, pick a subsequence of $\{f_n\}$ whose k -th derivatives converge to 0 in $L^q(d\mu)$ and use the same notation for that subsequence. Note that $|1 - (1 - \delta_n)z\bar{\zeta}_n| \approx \delta_n$ for $z \in SI_n$ and n large. Thus, by (2.5), we have

$$\begin{aligned} \|f_n^{(k)}\|_{L^q(d\mu)}^q &\geq \int_{SI_n} |f_n^{(k)}|^q d\mu \\ &\gtrsim \delta_n^{Nq-q(\alpha+2)/p} \int_{SI_n} |1 - (1 - \delta_n)z\bar{\zeta}_n|^{-Nq-kq} d\mu(z) \\ &\gtrsim C_2 \end{aligned}$$

for all large n . This is a contradiction, because $\|f_n^{(k)}\|_{L^q(d\mu)}^q \rightarrow 0$. The proof is complete. \square

Now, a change of variables and standard arguments give us the following Carleson measure characterizations of boundedness and compactness. As discussed in the first paragraph of this section, we restrict our consideration of the orders of differentiation to certain ranges; analysis of the general case seems too complicated for this paper. We also mention again that when $sp < \alpha + 1$ or $p = 2$, other methods are available and much more is known; see the discussion following Theorem 1.5 in the Introduction.

Theorem 2.6. *Assume that one of the following three conditions holds:*

- (i) $\alpha_1 > -1, 0 < p_1 \leq p_2$;
- (ii) $\alpha_1 = -1, p_1 = p_2 \geq 2$;
- (iii) $\alpha_1 = -1, 0 < p_1 < p_2$.

Also, assume $\alpha_2 > -1$ and

$$(2.6) \quad 0 \leq s_1 < 1 + \frac{\alpha_1 + 2}{p_1} - \frac{1}{p_2}, \quad 0 \leq s_2 < 1 + \frac{\alpha_2 + 1}{p_2}.$$

$$(a) \quad C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2} \text{ is bounded if and only if}$$

$$(2.7) \quad \mu_{p_2, \alpha_2 + (1-s_2)p_2}^\varphi(SI) = O\left(|I|^{(2+\alpha_1)p_2/p_1 + (1-s_1)p_2}\right), \quad I \subset \partial\mathbf{D}.$$

$$(b) \quad C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2} \text{ is compact if and only if}$$

$$(2.8) \quad \mu_{p_2, \alpha_2 + (1-s_2)p_2}^\varphi(SI) = o\left(|I|^{(2+\alpha_1)p_2/p_1 + (1-s_1)p_2}\right), \quad |I| \rightarrow 0.$$

Proof. Here, for brevity, we prove the sufficiency for boundedness and the necessity for compactness. The other implications can be seen by easy modifications. Also, let $\mu = \mu_{p_2, \alpha_2 + (1-s_2)p_2}^\varphi$ for simplicity.

First, we prove the sufficiency for boundedness. One may easily modify the proof for compactness. So, suppose that μ satisfies (2.7).

Note that

$$[p_2 - 2 + (2 + \alpha_1)p_2/p_1] - s_1p_2 > -1$$

by the first part of (2.6). Thus, by Lemma 2.5 (a) ($k = 0$), we have

$$(2.9) \quad \int_{\mathbf{D}} |g(z)|^{p_2} d\mu(z) \lesssim \int_{\mathbf{D}} |g(w)|^{p_2} (1 - |w|)^{[p_2 - 2 + (2 + \alpha_1)p_2/p_1] - s_1p_2} dA(w)$$

for functions g holomorphic on \mathbf{D} .

Also, note that $\alpha_2 + (1 - s_2)p_2 > -1$ by the second part of (2.6). It follows from (1.1), Proposition 2.2, Proposition 2.1 and (2.9) that

$$\begin{aligned} \|f \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}} &\approx \|f \circ \varphi\|_{A_{\alpha_2 + (1-s_1)p_2, 1}^{p_2}} \\ &\approx |f(\varphi(0))| + \|f'\|_{L^{p_2}(\mu)} \\ &\lesssim |f(\varphi(0))| + \|f'\|_{A_{[p_2 - 2 + (2 + \alpha_1)p_2/p_1] - s_1p_2}^{p_2}}. \end{aligned}$$

Now, by Proposition 2.2 and (1.1) again, we see that the sum in the last line above is equivalent to

$$\begin{aligned} \|f\|_{A_{[p_2 - 2 + (2 + \alpha_1)p_2/p_1] - s_1p_2, 1}^{p_2}} &\approx \|f\|_{A_{p_2 - 2 + (2 + \alpha_1)p_2/p_1, 1 + s_1}^{p_2}} \\ &= \|\mathcal{R}^{s_1} f\|_{A_{p_2 - 2 + (2 + \alpha_1)p_2/p_1, 1}^{p_2}} \\ &\approx |\mathcal{R}^{s_1} f(0)| + \|\partial \mathcal{R}^{s_1} f\|_{A_{p_2 - 2 + (2 + \alpha_1)p_2/p_1}^{p_2}}. \end{aligned}$$

Next, it is clear that $|\mathcal{R}^{s_1} f(0)| \lesssim \|f\|_{A_{\alpha_1, s_1}^{p_1}}$. Also, it is easy to verify using Lemma 2.5 (a) ($k = 1$) that

$$\|\partial \mathcal{R}^{s_1} f\|_{A_{p_2 - 2 + (2 + \alpha_1)p_2/p_1}^{p_2}} \lesssim \|\mathcal{R}^{s_1} f\|_{A_{\alpha_1}^{p_1}} = \|f\|_{A_{\alpha_1, s_1}^{p_1}}.$$

Putting these estimates together, we conclude the boundedness of $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$.

Next, we prove the necessity for compactness. So, suppose that $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is compact. Suppose that (2.8) does not hold. Then there exist a constant $C > 0$ and a sequence of arcs $I_n \subset \partial\mathbf{D}$ such that $|I_n| \rightarrow 0$ and

$$\mu(SI_n) \geq C|I_n|^{(2+\alpha_1)p_2/p_1 + (1-s_1)p_2}.$$

Let $\delta_n = |I_n|$ and $\zeta_n \in \partial\mathbf{D}$ be the center of I_n for each n . Fix a large integer $N > (\alpha_1 + 2)/p - s_1$. Let $g_n(z) = (1 - (1 - \delta_n)z\bar{\zeta}_n)^{-N}$ and put $f_n = g_n \|g_n\|_{A_{\alpha_1, s_1}^{p_1}}^{-1}$. Note $\|g_n\|_{A_{\alpha_1, s_1}^{p_1}} \approx \delta_n^{-N - s_1 + (2 + \alpha_1)/p_1}$ by Lemma 2.4. Thus, $\{f_n\}$ converges uniformly to 0 on compact subsets of \mathbf{D} . Therefore, using the compactness of $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$, we may pick a subsequence of $\{f_n \circ \varphi\}$ that converges to 0 in $A_{\alpha_2, s_2}^{p_2}$ and

use the same notation for that subsequence. Now, first using Proposition 2.1 and then proceeding as in the proof of Lemma 2.5, we have

$$\begin{aligned} \|f_n \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}} &\approx \|f_n \circ \varphi\|_{A_{\alpha_2 + (1-s_2)p_2, 1}^{p_2}} \\ &\gtrsim \delta_n^{N+s_1-(2+\alpha_1)/p_1} \left\{ \int_{SI_n} |1 - (1 - \delta_n)z\bar{\zeta}_n|^{-(N+1)p_2} d\mu(z) \right\}^{1/p_2} \\ &\gtrsim \delta_n^{s_1-1-(2+\alpha_1)/p_1} \mu(SI_n)^{1/p_2} \\ &\geq C \end{aligned}$$

for all large n . This is a contradiction, because $\|f_n \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}} \rightarrow 0$. The proof is complete. \square

3. FROM SMALL SPACES TO LARGER SPACES

We now turn to the proof of Theorem 1.3. For convenience we divide the theorem into more easily managed pieces, considering each implication separately as well as boundedness and compactness.

Theorem 3.1. *Let p_j , s_j and α_j ($j = 1, 2$) be as in the hypotheses of Theorem 2.6. In addition, assume that $\frac{\alpha_1+2}{p_1} - \frac{\alpha_2+2}{p_2} = s_1 - s_2$.*

- (a) $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded (compact, resp.) if and only if C_φ is bounded (compact, resp.) on $A_{\alpha_2, s_2}^{p_2}$.
- (b) If C_φ is bounded (compact, resp.) on $A_{\alpha_1, s_1}^{p_1}$, then so is C_φ on $A_{\alpha_2, s_2}^{p_2}$.

Proof. Note that $(2 + \alpha_1)p_2/p_1 + (1 - s_1)p_2 = \alpha_2 + 2 + (1 - s_2)p_2$. Thus, (a) follows from Theorem 2.6. Also, note that $A_{\alpha_1, s_1}^{p_1} \subset A_{\alpha_2, s_2}^{p_2}$ by (1.4). Thus, (b) follows from (a). \square

It is straightforward to check that when $\alpha_1 = -1$ and $p_1 = p_2 \geq 2$, the hypotheses (2.6) in Theorem 2.6 are equivalent to $s_2 < 1$ in (1.3). Thus, Theorem 1.3 with inclusion (1.3) is an immediate consequence of Theorem 3.1. Similarly, when $\alpha_1 \geq -1$ and $0 < p_1 < p_2$, the hypotheses (2.6) in Theorem 2.6 are equivalent to $\alpha_2 > -1$ and $s_2 < 1 + (1 + \alpha_2)/p_2$ in (1.4), and so Theorem 1.3 with inclusion (1.4) follows.

The proof of the next theorem uses some properties of the pseudo-hyperbolic distance ρ on \mathbf{D} . Recall that the pseudo-hyperbolic distance between points a and b in \mathbf{D} is given by

$$\rho(a, b) = \left| \frac{a - b}{1 - \bar{a}b} \right|.$$

We use $D(a, r)$ to denote the pseudo-hyperbolic disk of radius r and center a . Recall also the well-known and useful identity

$$1 - \left| \frac{a - b}{1 - \bar{a}b} \right|^2 = \frac{(1 - |a|^2)(1 - |b|^2)}{|1 - \bar{a}b|^2}, \quad a, b \in \mathbf{D}.$$

In particular, it is a consequence of this that

$$(3.1) \quad |1 - \bar{a}b| \approx 1 - |a|^2 \approx 1 - |b|^2$$

whenever $b \in D(a, 1/2)$.

The next result covers the inclusion (1.5) in Theorem 1.3, and so completes its proof.

Theorem 3.2. *Let $p > 0$, $\alpha \geq -1$ and $\frac{\alpha+2}{p} < s < 1 + \frac{\alpha+2}{p}$. If C_φ is bounded (compact, resp.) on $A^p_{\alpha,s}$, then so is C_φ on $\Lambda_{s-(\alpha+2)/p}$.*

Proof. We first prove the assertion on boundedness with the additional assumptions that $\alpha > -1$, $p > 1$ and $\frac{\alpha+2}{p} < s < 1 + \frac{\alpha+1}{p}$. Note that $\alpha + (1-s)p > -1$ and therefore $A^p_{\alpha,s} \approx A^p_{\alpha+(1-s)p,1}$ by (1.1). Choose $a \in \mathbf{D}$ such that $|\varphi(a)| \geq 1/2$, and consider the test function $f_a(z) = \log(1 - \overline{\varphi(a)}z)$. Then, by Proposition 2.2, we have

$$\begin{aligned} \|f_a \circ \varphi\|^p_{A^p_{\alpha+(1-s)p,1}} &\gtrsim \int_{\mathbf{D}} \frac{|\varphi(a)|^p}{|1 - \overline{\varphi(a)}\varphi(z)|^p} |\varphi'(z)|^p (1 - |z|^2)^{\alpha+(1-s)p} dA(z) \\ &\geq \int_{D(a,1/2)} \frac{|\varphi(a)|^p}{|1 - \overline{\varphi(a)}\varphi(z)|^p} |\varphi'(z)|^p (1 - |z|^2)^{\alpha+(1-s)p} dA(z). \end{aligned}$$

For $z \in D(a, 1/2)$, we have $1 - |z|^2 \approx 1 - |a|^2$, by (3.1). Also, the Schwarz-Pick Lemma tells us that $\varphi(z) \in D(\varphi(a), 1/2)$, and so $|1 - \overline{\varphi(a)}\varphi(z)| \approx 1 - |\varphi(a)|^2$ from (3.1). Using these estimates in the last term in the display above shows that

$$\begin{aligned} \|f_a \circ \varphi\|^p_{A^p_{\alpha+(1-s)p,1}} &\gtrsim \int_{D(a,1/2)} \frac{|\varphi'(z)|^p}{(1 - |\varphi(a)|^2)^p} (1 - |a|^2)^{\alpha+(1-s)p} dA(z) \\ (3.2) \qquad \qquad \qquad &\gtrsim \frac{|\varphi'(a)|^p (1 - |a|^2)^{\alpha+2+(1-s)p}}{(1 - |\varphi(a)|^2)^p}. \end{aligned}$$

For the last inequality we used that $D(a, 1/2)$ contains a Euclidean disk with center a and radius comparable to $1 - |a|^2$, and that $|\varphi'|^p$ is subharmonic.

Meanwhile, since $f(0) = 0$, we have

$$\begin{aligned} \|f_a\|^p_{A^p_{\alpha+(1-s)p,1}} &\approx \int_{\mathbf{D}} \frac{|\varphi(a)|^p}{|1 - \overline{\varphi(a)}z|^p} (1 - |z|^2)^{\alpha+(1-s)p} dA(z) \\ (3.3) \qquad \qquad \qquad &\approx (1 - |\varphi(a)|^2)^{\alpha+2-sp}, \end{aligned}$$

where the last equivalence holds by (2.3), because $sp > \alpha + 2$. Putting these estimates together with the assumption that $C_\varphi : A^p_{\alpha,s} \rightarrow A^p_{\alpha,s}$ is bounded, we get

$$\sup_{a \in \mathbf{D}} |\varphi'(a)| \left\{ \frac{(1 - |\varphi(a)|^2)}{(1 - |a|^2)} \right\}^{-1+s-(\alpha+2)/p} < \infty.$$

This is equivalent to the boundedness of C_φ on $\Lambda_{s-(\alpha+2)/p}$; see [Ma] or Theorem 4.9 in [CM].

Now, consider the general case $\alpha \geq -1$ and $\frac{\alpha+2}{p} < s < 1 + \frac{\alpha+2}{p}$. Choose $q > p$ so large that $q > 1$ and $s < 1 + \frac{\alpha+2}{p} - \frac{1}{q}$. Put $\beta = \frac{(\alpha+2)q}{p} - 2$. Then, $\beta > \alpha \geq -1$ and $\frac{\beta+2}{q} = \frac{\alpha+2}{p}$. Now, by (1.4) and (1.5), we have

$$A^p_{\alpha,s} \subset A^q_{\beta,s} \subset \Lambda_{s-(\beta+2)/q} = \Lambda_{s-(\alpha+2)/p}.$$

Also, note that $\frac{\beta+2}{q} < s < 1 + \frac{\beta+1}{q}$. Now, suppose that $C_\varphi : A^p_{\alpha,s} \rightarrow A^p_{\alpha,s}$ is bounded. Then $C_\varphi : A^q_{\beta,s} \rightarrow A^q_{\beta,s}$ is bounded by Theorem 3.1 and thus so is $C_\varphi : \Lambda_{s-(\alpha+2)/p} \rightarrow \Lambda_{s-(\alpha+2)/p}$ by the result for the special case we proved first. This proves the assertion on boundedness.

We now prove the assertion on compactness. Note that $\Lambda_{s-(\alpha+2)/p}$ and $A^p_{\alpha,s}$ are Möbius invariant, in the sense that every composition operator induced by a conformal automorphism of the unit disk maps each space into itself, and contained

in the disk algebra of holomorphic functions on the unit disk that extend to be continuous on the closed disk. Thus a general theorem of J. H. Shapiro [Sh] asserts that compactness of C_φ on each of these spaces implies that $\varphi(\mathbf{D})$ is a relatively compact subset of \mathbf{D} . We recall also that

$$|f(0)| + \sup\{(1 - |z|^2)^{1-\beta}|f'(z)| : z \in \mathbf{D}\}$$

is an equivalent norm on Λ_β ; see Theorem 4.1 in [CM]. Now, let $\{f_n\}$ be a bounded sequence in $\Lambda_{s-(\alpha+2)/p}$. We must show that some subsequence of $\{f_n \circ \varphi\}$ converges in $\Lambda_{s-(\alpha+2)/p}$. We know that $\{f_n\}$ is a normal family, and thus a subsequence (which we still call $\{f_n\}$) converges to some $f \in H(\mathbf{D})$ uniformly on compact subsets of \mathbf{D} . Also, if C_φ is compact on $A_{\alpha,s}^p$, then it is bounded and so $\varphi = C_\varphi z \in A_{\alpha,s}^p \subset \Lambda_{s-(\alpha+2)/p}$. Hence $(1 - |z|^2)^{1-s+(\alpha+2)/p}|\varphi'(z)|$ is uniformly bounded on \mathbf{D} , and it follows that

$$|f_n \circ \varphi(0) - f \circ \varphi(0)| + (1 - |z|^2)^{1-s+(\alpha+2)/p}|f'_n \circ \varphi(z) - f' \circ \varphi(z)||\varphi'(z)| \rightarrow 0$$

uniformly on \mathbf{D} as $n \rightarrow \infty$, since $\varphi(\mathbf{D})$ is contained in a compact subset of \mathbf{D} . This means that $\{f_n \circ \varphi\}$ converges to the function $g = f \circ \varphi$ in $\Lambda_{s-(\alpha+2)/p}$, and so $C_\varphi : \Lambda_{s-(\alpha+2)/p} \rightarrow \Lambda_{s-(\alpha+2)/p}$ is compact. The proof is complete. \square

Criteria for C_φ to be bounded or compact on Λ_ε are known. So Theorem 3.2 can be used to provide necessary conditions for boundedness or compactness of C_φ on the smaller spaces. In particular, we recall that the boundedness on Λ_ε implies the existence of the angular derivative of φ at all points of the unit circle where φ has a radial limit of modulus 1; see Corollary 4.10 in [CM]. This proves the following theorem.

Theorem 3.3. *Let $p > 0$, $\alpha \geq -1$ and $\frac{\alpha+2}{p} < s < 1 + \frac{\alpha+2}{p}$. If $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is bounded, then the angular derivative of φ exists at all points $\zeta \in \partial\mathbf{D}$ where the radial limit $\varphi(\zeta)$ of φ exists and satisfies $|\varphi(\zeta)| = 1$.*

As mentioned in the introduction, the conclusion of Theorem 3.3 is false for $\alpha = -1, s = 1$ and $p = 1$. Thus, for $\alpha = -1$, the lower bound $1/p$ for s cannot be decreased in general. We also give an example which shows that the lower bound $s > \frac{\alpha+2}{p}$ in Theorem 3.3 is sharp in case $\alpha > -1$. See Example 6.3 below.

The proof of the next theorem is based on Theorem 2.6. So, for simplicity, we restrict our consideration to the orders of differentiation covered there.

Theorem 3.4. *Let $p > 0$, $\alpha > -1$, $s \geq 0$ and assume $s < 1 + \frac{\alpha+1}{p}$.*

- (a) *If $1 + \frac{\alpha+1}{p} - s > \varepsilon \geq 0$ and C_φ is bounded (compact, resp.) on $A_{\alpha,s+\varepsilon}^p$, then so is C_φ on $A_{\alpha,s}^p$.*
- (b) *If $\varepsilon \geq 0$ and C_φ is bounded (compact, resp.) on $A_{\alpha,s}^p$, then so is C_φ on $A_{\alpha+\varepsilon,s}^p$.*

Proof. Let I be an arc in the unit circle, and let $\varphi(z) = w \in SI$. A standard argument using the Schwarz Lemma then tells us that $1 - |z| \lesssim 1 - |w| \leq |I|$, and so

$$\begin{aligned} N_{p,\alpha+\varepsilon+(1-s)p}(\varphi, w) &= \sum |\varphi'(z)|^{p-2} (1 - |z|^2)^{\alpha+\varepsilon+(1-s)p} \\ &\lesssim |I|^\varepsilon N_{p,\alpha+(1-s)p}(\varphi, w), \end{aligned}$$

$w \in SI$. Hence

$$\mu_{p,\alpha+\varepsilon+(1-s)p}^\varphi(SI) \lesssim |I|^\varepsilon \mu_{p,\alpha+(1-s)p}^\varphi(SI),$$

and statement (b) is now an immediate consequence of Theorem 2.6. The proof of (a) is similar and will be omitted. \square

We finish this section by giving the proof of Theorem 1.4 from the introduction, which we restate for convenience.

Theorem 3.5. *Let $p > 0$, $\alpha \geq -1$, $s \geq 0$ and assume $ps = \alpha + 2$.*

- (a) *If $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is bounded, then $C_\varphi : \text{VMOA} \rightarrow \text{VMOA}$ is bounded.*
- (b) *If φ is univalent and $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is compact, then $C_\varphi : \text{VMOA} \rightarrow \text{VMOA}$ is compact.*

In the proof below and elsewhere, we use the notation $\text{dist}(a, \partial E)$ for the Euclidean distance between a point a and the boundary of a set E .

Proof. If C_φ is bounded on $A_{\alpha,s}^p$, then from (1.6) we have that $\varphi = C_\varphi z \in A_{\alpha,s}^p \subset \text{VMOA}$. Also, it is easy to see that C_φ is bounded on VMOA if and only if $\varphi \in \text{VMOA}$; see, for example, [Sm2]. This gives part (a).

For the proof of (b), we recall that when φ is univalent, C_φ is compact on VMOA if and only if

$$(3.4) \quad \lim_{|w| \rightarrow 1^-} \frac{\text{dist}(w, \partial\varphi(\mathbf{D}))\chi_{\varphi(\mathbf{D})}(w)}{(1 - |w|)} = 0;$$

see Theorem 4.1 in [Sm2]. Also, it is an easy consequence of the Koebe distortion theorem that if φ is univalent, then

$$(3.5) \quad (1 - |z|^2)|\varphi'(z)| \approx \text{dist}(\varphi(z), \partial\varphi(\mathbf{D})), \quad z \in \mathbf{D};$$

see Corollary 1.4 in [P].

First, consider the case $p > 1$ and $\alpha > -1$. With $ps = \alpha + 2 < p + \alpha + 1$, case (i) of Theorem 2.6 (b) tells us that C_φ is compact on $A_{\alpha,s}^p$ if and only if $\mu_{p,p-2}^\varphi(SI) = o(|I|^p)$ as $|I| \rightarrow 0$. We prove part (b) by showing that this fails when C_φ is not compact on VMOA . From (3.4), if C_φ is not compact on VMOA , then there is an $\varepsilon > 0$ and a sequence $\{w_n\} \subset \varphi(\mathbf{D})$ with $|w_n| \rightarrow 1$ and $\text{dist}(w_n, \partial\varphi(\mathbf{D})) \geq \varepsilon(1 - |w_n|)$. Let I_n be the arc of the unit circle with center $w_n/|w_n|$ and length $|I_n| = 2(1 - |w_n|)$. Since φ is univalent,

$$\mu_{p,p-2}^\varphi(SI_n) = \int_{SI_n} \{|\varphi'(z)|(1 - |z|^2)\}^{p-2} dA(w),$$

where $w = \varphi(z)$. From (3.5), $|\varphi'(z)|(1 - |z|^2) \approx (1 - |w_n|)$ for w in the disk with center w_n and radius $\varepsilon(1 - |w_n|)/2$, which yields the lower bound

$$\mu_{p,p-2}^\varphi(SI_n) \gtrsim |I_n|^p.$$

Hence $\mu_{p,p-2}^\varphi(SI) \neq o(|I|^p)$, $|I| \rightarrow 0$, as desired.

Now, consider the general case $p > 0$, $\alpha \geq -1$ and suppose that $C_\varphi : A_{\alpha,s}^p \rightarrow A_{\alpha,s}^p$ is compact. With $ps = \alpha + 2$, choose q as in the proof of Theorem 3.2. That is, choose $q > p$ so large that $q > 1$ and put $\beta = sq - 2 > -1$. Then $\frac{\beta+2}{q} = \frac{\alpha+2}{p} = s$, and so $A_{\alpha,s}^p \subset A_{\beta,s}^q$ by (1.4). Thus, from Theorem 3.1 we see that $C_\varphi : A_{\beta,s}^q \rightarrow A_{\beta,s}^q$ is compact and thus so is $C_\varphi : \text{VMOA} \rightarrow \text{VMOA}$ by the result for the special case that we have proved above. The proof is complete. \square

4. COMPOSITION OPERATORS ON $A_{\alpha,s}^p$

In this section we prove Theorems 1.1 and 1.2 from the introduction. For convenience, we divide these results into more easily managed pieces. As mentioned in the introduction, it is well known that every composition operator is bounded on A_β^p for all $p > 0$ and $\beta > -1$. Note that we have $A_{\alpha,s}^p \approx A_{\alpha-sp}^p$ by (1.1), when $sp < \alpha + 1$. Thus, it follows that every composition operator is bounded on $A_{\alpha,s}^p$ whenever $sp < \alpha + 1$. The next two theorems complete the description of the general situation, as stated in Theorem 1.1.

Theorem 4.1. *Let $p > 0$, $s \geq 0$ and $\alpha \geq -1$. If $s = \frac{\alpha+1}{p}$ and*

- (a) $p \geq 2$ or $\alpha = -1$, *then every composition operator is bounded on $A_{\alpha,s}^p$;*
- (b) $p < 2$ and $\alpha > -1$, *then some composition operators are not bounded on $A_{\alpha,s}^p$.*

Proof. If $\alpha = -1$, then $s = 0$ and so every composition operator is bounded on $A_{\alpha,s}^p = H^p$. If $\alpha > -1$, $p \geq 2$ and $ps = \alpha + 1$, then from (1.3) we have that $H^p \subset A_{\alpha,s}^p$. Hence part (a) follows from Theorem 1.3, since all composition operators are bounded on H^p .

Turning to the proof of (b), first note that $A_{\alpha,s}^p \approx A_{p-1,1}^p$ by (1.1), since $s = \frac{\alpha+1}{p}$. Also, $\varphi = C_\varphi z \in A_{p-1,1}^p$ is necessary for C_φ to be bounded on $A_{p-1,1}^p$. Thus it suffices to show that if $p < 2$ there is a bounded analytic function $F \notin A_{p-1,1}^p$. The case $p = 1$ of this statement is outlined in exercise 9(a) in Chapter VI of [G]. That construction can be modified to work for $p < 2$. For completeness, we sketch the argument.

Let $p < 2$ and consider the function

$$f(z) = \sum_{k=1}^{\infty} k^{-1/p} z^{2^k}.$$

Since the series for f is lacunary with square summable coefficients, it is known that $f \in \text{BMOA}$. This is an easy consequence of BMOA being the dual of H^1 together with Paley's Inequality for the coefficients of an H^1 function ([D], p. 104), or see [Mi] for another approach to the proof. Next, it is easy to verify that if

$$z \in A_n = \{w \in \mathbf{D} : 1 - 2^{-n} \leq |w| < 1 - 2^{-n-1}\},$$

then $|f'(z)| \approx n^{-1/p} 2^n$. This leads to the approximation

$$\int_{A_n} |f'(z)|^p (1 - |z|^2)^{p-1} dA(z) \approx \frac{1}{n},$$

from which we see that $f \notin A_{p-1,1}^p$. This is not the required example, however, since f is not bounded. But since $f \in \text{BMOA}$, there are bounded functions u_1 and u_2 on the unit circle such that $\text{Re} f = u_1 + \tilde{u}_2$ where \tilde{u}_2 denotes the harmonic conjugate of u_2 . Here, we are using the same notation for a boundary function and its harmonic extension. Then $|f'|^p \lesssim |\nabla u_1|^p + |\nabla u_2|^p$ by the Cauchy-Riemann equations, and it follows that there is a bounded real function u on the circle such that

$$\int_{\mathbf{D}} |\nabla u(z)|^p (1 - |z|^2)^{p-1} dA(z) = \infty.$$

Now let $F = \exp(u + i\tilde{u})$, so that F is a bounded analytic function satisfying $|F'| \approx |\nabla u|$. Thus $F \notin A_{p-1,1}^p$, and the proof is complete. \square

The proof of the next theorem, covering the case $s > \frac{\alpha+1}{p}$, requires two lemmas, which will also be used in the next section.

Lemma 4.2. *Let $p > 0$, $s \geq 0$, and $\alpha \geq -1$. Then the following inclusions hold:*

- (a) $A^p_{\alpha,s} \subset A^p_{p+\alpha+\varepsilon,1+s}$ for $\varepsilon > 0$;
- (b) $A^p_{p+\alpha-\varepsilon,1+s} \subset A^p_{\alpha,s}$ for $0 < \varepsilon < p$.

Moreover, both inclusions are bounded.

We remark in passing that, when $s = \varepsilon = 0$, $\alpha = -1$ and $p \geq 2$, the inclusion in (a) holds and this is just a restatement of the well-known Littlewood-Paley inequality. When $s = \varepsilon = 0$, $\alpha = -1$ and $1 < p \leq 2$, the inclusion in (b) holds and this is a restatement of the dual of the Littlewood-Paley inequality.

Proof. By definition of the holomorphic Sobolev spaces, it is sufficient to consider the case $s = 0$. First, consider the case $\alpha > -1$. Then, we have

$$A^p_\alpha \subset A^p_{\alpha+\varepsilon} \approx A^p_{p+\alpha+\varepsilon,1}, \qquad \varepsilon > 0$$

and

$$A^p_{p+\alpha-\varepsilon,1} \approx A^p_{\alpha,\varepsilon/p} \subset A^p_{\alpha+\varepsilon,\varepsilon/p} \approx A^p_\alpha, \qquad 0 < \varepsilon < p$$

where the equivalences are from (1.1) and the inclusions are clearly bounded.

Now, assume $\alpha = -1$. Let $f \in H^p$ and put

$$(4.1) \qquad M^p_p(f,r) = \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta, \qquad 0 < r < 1.$$

Then, for any $\delta > 0$, we have $M_p(f,r) = O\left((1-r)^{-\delta}\right)$, which implies $M_p(f',r) = O\left((1-r)^{-1-\delta}\right)$ (see Theorem 5.5 of [D]). Thus, for $\varepsilon > 0$, integration using polar coordinates shows that

$$\int_{\mathbf{D}} |f'(z)|^p (1-|z|^2)^{p-1+\varepsilon} dA(z) \lesssim \int_0^1 (1-r)^{-(1+\delta)p+p-1+\varepsilon} dr.$$

With δ small enough so that $p\delta < \varepsilon$, this integral is convergent, and so (a) holds. Now, assume $0 < \varepsilon < p$. Consider the case $p > 1$ first. Let p' be the conjugate exponent of p . By the fundamental theorem of calculus and Hölder's inequality, we have

$$\begin{aligned} \int_0^{2\pi} |f(e^{i\theta}) - f(0)|^p d\theta &\leq \int_0^{2\pi} \left\{ \int_0^1 |f'(re^{i\theta})| dr \right\}^p d\theta \\ &\leq C \int_0^{2\pi} \left\{ \int_0^1 |f'(re^{i\theta})|^p (1-r)^{(p-1-\varepsilon)} dr \right\} d\theta \\ &\approx \int_{\mathbf{D}} |f'(z)|^p (1-|z|^2)^{p-1-\varepsilon} dA(z), \end{aligned}$$

where

$$C = \left\{ \int_0^1 (1-r)^{-p'(p-1-\varepsilon)/p} dr \right\}^{p/p'} < \infty.$$

It follows that

$$\|f\|_{H^p}^p \lesssim |f(0)|^p + \int_{\mathbf{D}} |f'(z)|^p (1-|z|^2)^{p-1-\varepsilon} dA(z),$$

and the same is true for $p = 1$ by a trivial modification of this argument. This proves (b) for $p \geq 1$ by Proposition 2.2. When $0 < p < 1$, we have the inclusions

$$A_{p-1-\varepsilon,1}^p \subset A_{-1,0}^{p/(1-\varepsilon)} = H^{p/(1-\varepsilon)} \subset H^p$$

where the first inclusion comes from (1.4). This completes the proof. \square

We next show that certain inclusions between holomorphic Sobolev spaces are compact.

Lemma 4.3. *Let $p > 0$, $s \geq 0$, $\varepsilon > 0$ and $\alpha \geq -1$. Then the following inclusions are compact:*

$$A_{\alpha,s+\varepsilon}^p \subset A_{\alpha,s}^p \subset A_{\alpha+\varepsilon,s}^p.$$

Proof. We first consider the case $\alpha > -1$. Let $\{f_n\}$ be a bounded sequence in $A_{\alpha,s}^p$. To show that the inclusion $A_{\alpha,s}^p \subset A_{\alpha+\varepsilon,s}^p$ is compact, we must show that some subsequence of $\{f_n\}$ converges in $A_{\alpha+\varepsilon,s}^p$. It is well known that the inclusion $A_{\alpha}^p \subset A_{\alpha+\varepsilon}^p$ is compact; see, for example, [Sm1]. Thus there exists $g \in A_{\alpha+\varepsilon}^p$ and a subsequence of $\{f_n\}$ (which for convenience we continue to denote $\{f_n\}$) such that $\mathcal{R}^s f_n \rightarrow g$ in $A_{\alpha+\varepsilon}^p$. Now, choose $h \in H(\mathbf{D})$ such that $\mathcal{R}^s h = g$. It is then clear that $h \in A_{\alpha+\varepsilon,s}^p$ and $f_n \rightarrow h$ in $A_{\alpha+\varepsilon,s}^p$. This completes the proof that $A_{\alpha,s}^p \subset A_{\alpha+\varepsilon,s}^p$ is compact. Now, since we have by (1.1),

$$A_{\alpha,s+\varepsilon}^p \subset A_{\alpha+p\varepsilon,s+\varepsilon}^p \approx A_{\alpha,s}^p$$

and the first inclusion is compact, we conclude the compactness of the inclusion $A_{\alpha,s+\varepsilon}^p \subset A_{\alpha,s}^p$.

Now, assume $\alpha = -1$. Choose positive numbers δ_1, δ_2 such that $\delta_2 < \delta_1 < \min(p, p\varepsilon)$. Then, we have the relations

$$A_{-1,s+\varepsilon}^p \subset A_{p-1+p\varepsilon-\delta_1,1+s+\varepsilon}^p \approx A_{p-1-\delta_1,1+s}^p \subset A_{p-1-\delta_2,1+s}^p \subset A_{-1,s}^p.$$

The first and last inclusions are bounded from Lemma 4.2, the equivalence is from (1.1), while the remaining inclusion is compact by the previously established part of this lemma. Hence $A_{-1,s+\varepsilon}^p \subset A_{-1,s}^p$ is compact. For the compactness of $A_{-1,s}^p \subset A_{-1+\varepsilon,s}^p$ note that

$$A_{-1,s}^p \subset A_{p-1+\varepsilon/2,1+s}^p \approx A_{-1+\varepsilon/2,s}^p \subset A_{-1+\varepsilon,s}^p.$$

The first inclusion is bounded from Lemma 4.2, as the isomorphism from (1.1), while the remaining inclusion is compact by a previously established part of this lemma for $\alpha > -1$. Hence $A_{-1,s}^p \subset A_{-1+\varepsilon,s}^p$ is compact. The proof is complete. \square

Theorem 4.4. *Let $p > 0$ and $\alpha \geq -1$. If $s > \frac{\alpha+1}{p}$, then some composition operators are not bounded on $A_{\alpha,s}^p$.*

Proof. Since $\varphi = C_{\varphi}z \in A_{\alpha,s}^p$ is necessary for C_{φ} to be bounded on $A_{\alpha,s}^p$, it suffices to show that if $s > \frac{\alpha+1}{p}$, then $H^{\infty} \setminus A_{\alpha,s}^p \neq \emptyset$, where H^{∞} denotes the class of all bounded holomorphic functions on \mathbf{D} . Suppose to the contrary that $H^{\infty} \subset A_{\alpha,s}^p$. Then this inclusion map is continuous by the Closed Graph Theorem, while the inclusion map $A_{\alpha,s}^p \subset A_{\beta,s}^p$ was shown to be compact whenever $\alpha < \beta$ in Lemma 4.3. The hypothesized lower bound for s now allows us to choose $\delta \geq 1$ such that $\alpha + \delta < sp \leq \alpha + \delta + 1$. Moreover, we can choose $\varepsilon > 0$ so small that

$0 < \alpha + \delta + 1 + \varepsilon - sp < 1$, which gives $A_{\alpha+\delta-1+\varepsilon,s}^p \subset H^{p/(\alpha+\delta+1+\varepsilon-sp)}$ from (1.4). Consequently, we have a chain of inclusions

$$H^\infty \subset A_{\alpha,s}^p \subset A_{\alpha+\delta-1+\varepsilon,s}^p \subset H^{p/(\alpha+\delta+1+\varepsilon-sp)}.$$

Thus the inclusion $H^\infty \subset H^{p/(\alpha+\delta+1+\varepsilon-sp)}$ can be viewed as a product of a compact map and two bounded maps, and hence is compact. But $\{z^n\}$ is a bounded sequence in H^∞ for which no subsequence converges in $H^{p/(\alpha+\delta+1+\varepsilon-sp)}$. This contradicts the compactness of the inclusion map, and the proof is complete. \square

The next theorem (also stated as Theorem 1.2 in the introduction) shows that the upper bounds for s in Theorems 4.1 and 4.4 can be extended when the symbol φ of the composition operator is of bounded valence.

Theorem 4.5. *Let $\alpha > -1$ and let φ be of bounded valence. Assume that either $0 \leq s \leq \frac{\alpha+2}{p}$ if $p \geq 2$, or $0 \leq s < \frac{\alpha+1}{p} + \frac{1}{2}$ if $0 < p < 2$. Then C_φ is bounded on $A_{\alpha,s}^p$.*

Proof. We use the Carleson measure criteria from Theorem 2.6. We need to estimate

$$\mu_{p,\alpha+(1-s)p}^\varphi(SI) = \int_{SI} \sum_{\varphi(z)=w} |\varphi'(z)|^{p-2} (1-|z|^2)^{\alpha+(1-s)p} dA(w)$$

for arbitrary arcs $I \subset \partial\mathbf{D}$.

First, consider the case $p \geq 2$. By assumption we have $sp \leq \alpha + 2$. Note that, since φ is of bounded valence, there is a uniformly bounded number of terms in the sum inside the integral above. Next, we set $w = \varphi(z)$ and use the Schwarz-Pick Lemma, which asserts that $|\varphi'(z)| \leq (1-|w|^2)/(1-|z|^2)$, and then the elementary inequality $1-|z|^2 \leq C(1-|w|^2)$ to get that

$$\begin{aligned} \mu_{p,\alpha+(1-s)p}^\varphi(SI) &\leq \int_{SI} (1-|w|^2)^{p-2} \sum_{z \in \varphi^{-1}\{w\}} (1-|z|^2)^{\alpha+2-sp} dA(w) \\ &\lesssim \int_{SI} (1-|w|^2)^{\alpha+(1-s)p} dA(w) \\ &\lesssim |I|^{\alpha+2+(1-s)p}, \end{aligned}$$

which from Theorem 2.6 is equivalent to boundedness of C_φ on $A_{\alpha,s}^p$. We note that the hypothesis $p \geq 2$ was used in getting the first inequality, and $sp \leq \alpha + 2$ was used in the second inequality.

Now, consider the case $p < 2$. What we have now is $sp < \alpha + 1 + \frac{p}{2}$. By the area formula (Theorem 2.32 in [CM]), we have

$$\mu_{p,\alpha+(1-s)p}^\varphi(SI) = \int_{\varphi^{-1}(SI)} |\varphi'(z)|^p (1-|z|^2)^{\alpha+(1-s)p} dA(z).$$

Note that, since φ is of bounded valence, we have

$$\int_{\varphi^{-1}(SI)} |\varphi'(z)|^2 dA(z) = \int_{SI} \sum_{z \in \varphi^{-1}\{w\}} 1 dA(w) \lesssim |I|^2$$

and thus

$$\begin{aligned}
 \mu_{p,\alpha+(1-s)p}^\varphi(SI) &\leq \left(\int_{\varphi^{-1}(SI)} |\varphi'(z)|^2 dA(z) \right)^{p/2} \\
 &\quad \times \left(\int_{\varphi^{-1}(SI)} (1 - |z|^2)^{2(\alpha+(1-s)p)/(2-p)} dA(z) \right)^{(2-p)/2} \\
 (4.2) \quad &\lesssim |I|^p \left(\int_{\varphi^{-1}(SI)} (1 - |z|^2)^{2(\alpha+(1-s)p)/(2-p)} dA(z) \right)^{(2-p)/2}
 \end{aligned}$$

where the first inequality is provided by Hölder's inequality. To estimate the integral above, recall that every composition operator C_φ is bounded on A_β^p , $\beta > -1$, and that this is equivalent to

$$(4.3) \quad \int_{\varphi^{-1}(SI)} (1 - |z|^2)^\beta dA(z) \lesssim |I|^{\beta+2}$$

for all $I \subset \partial\mathbf{D}$; see section 4 in [MS]. Since, by hypothesis, $2(\alpha + (1-s)p)/(2-p) > -1$, we can combine the estimates in (4.2) and (4.3) to get that

$$\mu_{p,\alpha+(1-s)p}^\varphi(SI) \lesssim |I|^{p+\alpha+(1-s)p+(2-p)} = |I|^{\alpha+2+(1-s)p}.$$

From Theorem 2.6, this is equivalent to C_φ being bounded on $A_{\alpha,s}^p$. The proof is complete. \square

Our final result in this section shows that the compactness of the inclusions in Lemma 4.3 extends to all composition operators with a certain restriction on s .

Theorem 4.6. *Let $p > 0$, $\alpha \geq -1$ and $\varepsilon > 0$. Assume that either (i) $\alpha > -1$, $0 \leq s < \frac{\alpha+1}{p}$ or (ii) $\alpha = -1$, $s = 0$ or (iii) $s = \frac{\alpha+1}{p}$, $p \geq 2$.*

(a) $C_\varphi : A_{\alpha,s+\varepsilon}^p \rightarrow A_{\alpha,s}^p$ is compact.

(b) $C_\varphi : A_{\alpha-\varepsilon,s}^p \rightarrow A_{\alpha,s}^p$ is compact whenever $\alpha - \varepsilon \geq -1$.

Proof. We may view the action of $C_\varphi : A_{\alpha,s+\varepsilon}^p \rightarrow A_{\alpha,s}^p$ as follows:

$$A_{\alpha,s+\varepsilon}^p \subset A_{\alpha,s}^p \xrightarrow{C_\varphi} A_{\alpha,s}^p.$$

Also, if $\alpha - \varepsilon \geq -1$, then we may view the action of $C_\varphi : A_{\alpha-\varepsilon,s}^p \rightarrow A_{\alpha,s}^p$ as follows:

$$A_{\alpha-\varepsilon,s}^p \subset A_{\alpha,s}^p \xrightarrow{C_\varphi} A_{\alpha,s}^p.$$

The inclusions above are both compact by Lemma 4.3, and every composition operator is bounded on $A_{\alpha,s}^p$ by Theorem 1.1. Thus, both operators are compact. The proof is complete. \square

Carleson measure criteria for C_φ to be bounded or compact on H_1^p are known; see Theorems 4.11 and 4.12 in [CM]. For s not an integer, characterizing when C_φ is bounded or compact on H_s^p seems much harder than the analogous problems on the Bergman spaces. The problem is that for the Hardy spaces, (1.1) does not provide isomorphisms with spaces defined using full derivatives. Thus, we are led to the following.

Problem. Characterize φ for which C_φ is bounded (compact) on H_s^p , $s > 0$.

5. COMPOSITION WITH A POLYGONAL MAP

Here, we find simple criteria for the compactness of the composition operators between holomorphic Sobolev spaces induced by polygonal maps.

Recall that $\text{dist}(a, \partial E)$ denotes the Euclidean distance between a point a and the boundary of a set E .

Lemma 5.1. *Suppose that P is a polygon inscribed in the unit circle with a vertex at v and let $\pi\eta(v)$ be the vertex angle at v . Then, given a Riemann map φ of \mathbf{D} onto P , there exists a neighborhood N_v of v such that*

$$\begin{aligned} |\varphi'(z)| &\approx (1 - |\varphi(z)|)^{1-1/\eta(v)}, \\ (1 - |z|) &\approx (1 - |\varphi(z)|)^{1/\eta(v)-1} \text{dist}(\varphi(z), \partial P) \end{aligned}$$

for all $z \in \varphi^{-1}(N_v)$.

Proof. Recall that φ extends to a homeomorphism of $\overline{\mathbf{D}}$ onto \overline{P} (see, for example, Theorem 14.19 of [R2]). Assume $v = 1$ and $\varphi(1) = 1$ for simplicity. Also, let $\eta = \eta(v)$. Then, a reflection argument yields

$$1 - \varphi(z) = c(1 - z)^\eta + O(|1 - z|^{1+\eta})$$

for some constant $c \neq 0$ and for all z near 1. Thus, we have

$$|\varphi'(z)| \approx |1 - z|^{\eta-1} \approx |1 - \varphi(z)|^{1-1/\eta} \approx (1 - |\varphi(z)|)^{1-1/\eta}$$

for z near 1. The last equivalence in the display above holds, because $\varphi(z)$ is contained in a nontangential region with the vertex at 1. This proves the first equivalence of the lemma. The second equivalence is now a consequence of the estimates

$$\frac{1}{4}(1 - |z|^2)|\varphi'(z)| \leq \text{dist}(\varphi(z), \partial P) \leq (1 - |z|^2)|\varphi'(z)|,$$

which hold since φ is univalent; see Corollary 1.4 of [P]. The proof is complete. \square

Lemma 5.2. *Let P be a polygon inscribed in the unit circle. Assume $b > -1$ and $a + b > -2$. Then, there exists a constant $C > 0$ such that*

$$\int_{P \cap SI} (1 - |w|^2)^a \text{dist}(w, \partial P)^b dA(w) \leq C|I|^{a+b+2}$$

for all arcs $I \subset \partial \mathbf{D}$.

Proof. Let us introduce a temporary notation. For an arc $I \subset \partial \mathbf{D}$ with center at $\zeta \in \partial \mathbf{D}$ and $|I| = 2\delta$, we let $S_\delta(\zeta) = SI$.

Assume that δ is sufficiently small and $P \cap S_\delta(\zeta) \neq \emptyset$. Then, there is a constant C_1 , depending only on P , such that $S_\delta(\zeta) \subset S_{C_1\delta}(v)$ for some vertex v of P . Assume $v = 1$ for simplicity. Assuming that δ is sufficiently small so that $S_{C_1\delta}(1)$ contains no vertex of P other than 1, note that $1 - |w| \approx |1 - w|$ for $w \in P \cap S_{C_1\delta}(1)$. Now,

we have

$$\begin{aligned}
 & \int_{P \cap S_\delta(\zeta)} (1 - |w|^2)^a \operatorname{dist}(w, \partial P)^b dA(w) \\
 & \leq \int_{P \cap S_{C_1 \delta}(1)} (1 - |w|^2)^a \operatorname{dist}(w, \partial P)^b dA(w) \\
 & \approx \int_{P \cap S_{C_1 \delta}(1)} |1 - w|^a \operatorname{dist}(w, \partial P)^b dA(w) \\
 & \lesssim \int_0^\pi \int_0^{C_2 \delta} r^a (r \sin \theta)^b r dr d\theta \\
 & \approx \delta^{a+b+2}
 \end{aligned}$$

as asserted, where C_2 is a constant depending only on P . The estimate for large δ follows from the inequality

$$\int_P (1 - |w|^2)^a \operatorname{dist}(w, \partial P)^b dA(w) < \infty,$$

which is clear from the argument above. The proof is complete. \square

Recall that $D(z, 1/2)$ denotes the pseudohyperbolic disk. Let $D(z) = D(z, 1/2)$. In the following we let $dA_\alpha(z) = (1 - |z|^2)^\alpha dA(z)$ for $\alpha > -1$. The following lemma is proved for $\alpha = 0$ in [L2], and the same proof works for general α .

Lemma 5.3. *Let $\alpha > -1$ and μ be a positive finite Borel measure on \mathbf{D} . Assume $p > q > 0$. Then, there is a constant C such that*

$$\left(\int_{\mathbf{D}} |f|^q d\mu \right)^{1/q} \leq C \left(\int_{\mathbf{D}} |f|^p dA_\alpha \right)^{1/p}, \quad f \in A_\alpha^q$$

if and only if $\tau \in L^{\frac{p}{p-q}}(A_\alpha)$ where $\tau(z) = \frac{\mu(D(z))}{A_\alpha(D(z))}$.

For a polygon P inscribed in the unit circle, recall that $\theta(P)$ denotes $1/\pi$ times the measure of the largest vertex angle of P .

Proposition 5.4. *Let $p_j > 0$, $\alpha_j > -1$, $s_j \geq 0$ ($j = 1, 2$) and assume*

$$(5.1) \quad s_1 < 1 + \frac{\alpha_1 + 2}{p_1} - \frac{1}{p_2}, \quad s_2 < \frac{\alpha_2 + 1}{p_2}.$$

Let φ be a holomorphic function taking \mathbf{D} into a polygon P inscribed in the unit circle. If

$$(5.2) \quad \theta(P) < \frac{p_1(\alpha_2 + 2 - s_2 p_2)}{p_2(\alpha_1 + 2 - s_1 p_1)},$$

then $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded.

Moreover, for functions φ of bounded valence, the second part of (5.1) can be replaced by the weaker condition that

$$(5.3) \quad s_2 \leq \frac{\alpha_2 + 2}{p_2} \text{ if } p_2 \geq 2, \text{ or } s_2 < \frac{\alpha_2 + 1}{p_2} + \frac{1}{2} \text{ if } 0 < p_2 < 2.$$

In either case the equality can be allowed in (5.2) for $p_2 \geq p_1$.

Proof. Let φ_0 be a Riemann mapping of \mathbf{D} onto P and put $\psi = \varphi_0^{-1} \circ \varphi$. Then $\varphi = \varphi_0 \circ \psi$ and thus $C_\varphi = C_{\varphi_0 \circ \psi} = C_\psi C_{\varphi_0}$. Note that $C_\psi : A_{\alpha_2, s_2}^{p_2} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded by Theorem 1.1 or Theorem 1.2. This shows that we only need to prove the proposition for $\varphi = \varphi_0$. So, in the rest of the proof, we assume that φ is a Riemann map of \mathbf{D} onto P . For simplicity, let $\beta_j = p_j + \alpha_j - s_j p_j$ and let $\gamma_j = 2 + \alpha_j - s_j p_j$ for $j = 1, 2$.

First, consider the case $p_2 \geq p_1$. By Theorem 2.6, we need to show that

$$(5.4) \quad \mu_{p_2, \beta_2}^\varphi(SI) = O(|I|^{(2+\alpha_1)p_2/p_1+(1-s_1)p_2})$$

for all arcs I . As in the proof of Lemma 5.2, we only need to consider I centered at a vertex of P for which $|I|$ is sufficiently small. Given such I , we have by Lemma 5.1 and Lemma 5.2,

$$\begin{aligned} \mu_{p_2, \beta_2}^\varphi(SI) &= \int_{SI \cap P} |\varphi'(\varphi^{-1}(w))|^{p_2-2} (1 - |\varphi^{-1}(w)|^2)^{\beta_2} dA(w) \\ &\approx \int_{SI \cap P} \text{dist}(w, \partial P)^{\beta_2} (1 - |w|)^{(1/\theta-1)\gamma_2} dA(w) \\ (5.5) \quad &\lesssim |I|^{p_2+\gamma_2/\theta} \end{aligned}$$

where $\theta = \theta(P)$. In the last inequality we used the fact that $\beta_2 > -1$, $\gamma_2 \geq 0$ from (5.3) and thus

$$(5.6) \quad \beta_2 + \left(\frac{1}{\theta} - 1\right) \gamma_2 > -1.$$

Thus, we have (5.4) by (5.2) and (5.5). Also, the same proof works in case the equality holds in (5.2).

Next, consider the case $p_2 < p_1$. We may assume $\varphi(0) = 0$. Let $f \in A_{\alpha_1, s_1}^{p_1}$ be an arbitrary function such that $f(0) = 0$. Since $p_2 < p_1$, we have $\beta_1 > -1$ by the first part of (5.1). Thus, by (1.1) and Proposition 2.2, we have

$$\|f\|_{A_{\alpha_1, s_1}^{p_1}}^{p_1} \approx \int_{\mathbf{D}} |f'(w)|^{p_1} (1 - |w|)^{\beta_1} dA(w).$$

Also, we have by (1.1), Proposition 2.2 and Lemma 5.1,

$$\begin{aligned} \|C_\varphi(f)\|_{A_{\alpha_2, s_2}^{p_2}}^{p_2} &\approx \|C_\varphi(f)\|_{A_{\beta_2, 1}^{p_2}}^{p_2} \\ &\approx \int_{\mathbf{D}} |f'(\varphi(z))|^{p_2} |\varphi'(z)|^{p_2} (1 - |z|^2)^{\beta_2} dA(z) \\ &= \int_P |f'(w)|^{p_2} |\varphi'(\varphi^{-1}(w))|^{p_2-2} (1 - |\varphi^{-1}(w)|^2)^{\beta_2} dA(w) \\ &\approx \int_P |f'(w)|^{p_2} \text{dist}(w, \partial P)^{\beta_2} (1 - |w|)^{(1/\theta-1)\gamma_2} dA(w). \end{aligned}$$

Now, define measures

$$\begin{aligned} d\mu_1(w) &= (1 - |w|)^{\beta_1} dA(w), \\ d\mu_2(w) &= \text{dist}(w, \partial P)^{\beta_2} (1 - |w|)^{(1/\theta-1)\gamma_2} \chi_{\varphi(\mathbf{D})}(w) dA(w) \end{aligned}$$

and let

$$\tau(z) = \frac{\mu_2(D(z))}{\mu_1(D(z))}, \quad z \in \mathbf{D}.$$

By Lemma 5.3, we need to show that $\tau \in L^p(\mu_1)$ where $p = p_1/(p_1 - p_2)$. Note that $\mu_2(D(z)) = 0$ if z is outside of some polygonal region Q . On the other hand, for $z \in Q$, we have

$$\begin{aligned}\mu_1(D(z)) &\approx (1 - |z|^2)^{\beta_1+2}, \\ \mu_2(D(z)) &\approx (1 - |z|^2)^{\beta_2+(1/\theta-1)\gamma_2+2};\end{aligned}$$

the first estimate is standard and the second one can be verified with (5.6) by modifying the proof of Lemma 5.2. Accordingly, we have

$$\tau(z) \approx (1 - |z|^2)^{\beta_2-\beta_1+(1/\theta-1)\gamma_2} \chi_Q(z).$$

It follows that $\tau \in L^p(\mu_1)$ if and only if $p[\beta_2 - \beta_1 + (1/\theta - 1)\gamma_2] + \beta_1 > -2$, which turns out to be the same as (5.2). This completes the proof. \square

Theorem 5.5. *Let $p_j > 0$, $\alpha_j \geq -1$, $s_j \geq 0$ ($j = 1, 2$) and assume (5.1) holds. Let φ be a holomorphic function taking \mathbf{D} into a polygon P inscribed in the unit circle. If (5.2) holds, then $C_\varphi : A_{\alpha_1, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is compact.*

Moreover, for functions φ of bounded valence, the second part of (5.1) can be replaced by the weaker condition (5.3).

In Example 6.1 below, we show that (5.3) provides the sharp upper bound of $\frac{\alpha_2+2}{p_2}$ for s_2 when $p_2 \geq 2$. While we do not know whether it does the same when $p_2 < 2$, the upper bound for s_2 when $p_2 < 2$ cannot be extended to $\frac{\alpha_2+2}{p_2}$, as is shown by Example 6.2. Nevertheless, Example 6.4 shows the upper bound of $\theta(P)$ in (5.2) is sharp in either case.

Proof. Assume that (5.2) holds and choose $\varepsilon > 0$ sufficiently small so that (5.2) holds with $\alpha_1 + \varepsilon$ in place of α_1 . By Lemma 4.3 we have $A_{\alpha_1, s_1}^{p_1} \subset A_{\alpha_1+\varepsilon, s_1}^{p_1}$ and the inclusion is compact. Thus, it is sufficient to show that $C_\varphi : A_{\alpha_1+\varepsilon, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded. In case $\alpha_2 > -1$, we see that $C_\varphi : A_{\alpha_1+\varepsilon, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded by Proposition 5.4.

So, assume $\alpha_2 = -1$. Note that with $\alpha_2 = -1$ there is no s_2 satisfying the second part of (5.1). Thus, we only need to be concerned about the case where φ is of bounded valence and (5.3) holds. First, consider the case $p_2 \leq 2$. In this case, we can view the action of C_φ as follows:

$$A_{\alpha_1+\varepsilon, s_1}^{p_1} \xrightarrow{C_\varphi} A_{p_2-1, s_2+1}^{p_2} \subset A_{-1, s_2}^{p_2}$$

where $C_\varphi : A_{\alpha_1+\varepsilon, s_1}^{p_1} \rightarrow A_{p_2-1, s_2+1}^{p_2}$ is bounded by Proposition 5.4 and the inclusion comes from (1.2). Therefore, $C_\varphi : A_{\alpha_1+\varepsilon, s_1}^{p_1} \rightarrow A_{\alpha_2, s_2}^{p_2}$ is bounded. Next, consider the case $p_2 > 2$. Choose $p'_2 \in (2, p_2)$ and $\alpha'_2 > -1$. Also, let $s'_2 = \frac{\alpha'_2+2}{p'_2} + s_2 - \frac{1}{p_2}$. Then, we can view the action of C_φ as follows:

$$A_{\alpha_1+\varepsilon, s_1}^{p_1} \xrightarrow{C_\varphi} A_{\alpha'_2, s'_2}^{p'_2} \subset A_{-1, s_2}^{p_2}$$

where $C_\varphi : A_{\alpha_1+\varepsilon, s_1}^{p_1} \rightarrow A_{\alpha'_2, s'_2}^{p'_2}$ is bounded by Proposition 5.4 and the inclusion comes from (1.4). The proof is complete. \square

Remarks. 1. As mentioned in the proof above, there is no s_2 satisfying the second part of (5.1) in case $\alpha_2 = -1$. Thus, we have no conclusion in Theorem 5.5 for general φ in case the target space is a Hardy-Sobolev space.

2. Note that the condition (5.2) holds vacuously if $\frac{\alpha_1+2}{p_1} - \frac{\alpha_2+2}{p_2} \leq s_1 - s_2$.

6. EXAMPLES

We now give several examples demonstrating that our theorems are sharp. For that purpose we introduce the so-called lens maps. For $0 < \eta < 1$ we denote by φ_η the function defined by

(6.1)
$$\varphi_\eta(z) = \frac{\sigma(z)^\eta - 1}{\sigma(z)^\eta + 1}, \quad z \in \mathbf{D}$$

where $\sigma(z) = (1 + z)/(1 - z)$. Let $\varphi_\eta(\mathbf{D}) = L_\eta$. Then, φ_η is the Riemann map of \mathbf{D} onto the subset L_η of \mathbf{D} bounded by arcs of circles meeting at $z = \pm 1$ at an angle of $\eta\pi$, and fixing the points $-1, 0$, and 1 . Because of the shape of the range L_η , such a map is called a “lens map”. Note that L_η is contained in a polygon inscribed in the unit circle.

By a straightforward calculation, we have

(6.2)
$$1 - |\varphi(z)| \approx |1 - z|^\eta, \quad |\varphi'(z)| \approx |1 - z|^{\eta-1}$$

for z near 1 .

The first example shows that the upper bound $s \leq \frac{\alpha+2}{p}$ in Theorem 1.2(a) is sharp. Also, this example shows that the upper bound $s_2 \leq \frac{\alpha_2+2}{p_2}$ in Theorem 5.5 is sharp when $p_2 \geq 2$ and φ is of bounded valence.

Example 6.1. Let $p > 1$, $\alpha > -1$ and $\frac{\alpha+2}{p} < s < 1 + \frac{\alpha+1}{p}$. Then, there exists a lens map $\varphi_\eta \notin A^p_{\alpha,s}$. In particular, C_{φ_η} is not bounded on $A^p_{\alpha,s}$.

Proof. Choose $0 < \eta < 1$ sufficiently small so that $sp \geq \eta p + \alpha + 2$ and consider the corresponding lens map φ_η . Note that $A^p_{\alpha,s} \approx A^p_{\alpha+(1-s)p,1}$ by (1.1). Therefore, we have by Proposition 2.2 and (6.2),

$$\begin{aligned} \|\varphi_\eta\|_{A^p_{\alpha,s}} &\approx \int_{\mathbf{D}} |\varphi'_\eta(z)|^p (1 - |z|^2)^{\alpha+(1-s)p} dA(z) \\ &\approx \int_{\mathbf{D}} |1 - z|^{(\eta-1)p} (1 - |z|^2)^{\alpha+(1-s)p} dA(z). \end{aligned}$$

Note that $(\eta - 1)p + \alpha + (1 - s)p \leq -2$, because $sp \geq \eta p + \alpha + 2$. Thus, an obvious estimate in an angle with vertex at 1 shows that the last integral above diverges, as desired. □

We do not know whether the upper bound $s < \frac{\alpha+1}{p} + \frac{1}{2}$ in Theorem 1.2(b) is sharp. However, the next example shows that the upper bound cannot be extended to $\frac{\alpha+2}{p}$ as in Theorem 1.2(a). Also, this is related to the assumption, $s_2 < \frac{\alpha_2+1}{p_2} + \frac{1}{2}$, in Theorem 5.5 when $p_2 < 2$.

Example 6.2. For each $p \in [1, 2)$, there exist $\alpha > -1$, $0 \leq s < 1$ with $s < \frac{\alpha+2}{p}$ and a univalent holomorphic self-map φ of \mathbf{D} such that $\varphi \notin A^p_{\alpha,s}$. In particular, C_φ is not bounded on $A^p_{\alpha,s}$.

Proof. P. Jones and N. Makarov have shown (see Theorem D(2) in [JM]) that for any $p < 2$, there exist a univalent holomorphic self-map φ_p of \mathbf{D} and a constant $c > 0$ such that the integral means of φ'_p satisfy

(6.3)
$$(1 - r_n)^{p-1+c(2-p)^2} M^p_p(\varphi'_p, r_n) \geq 1$$

for some sequence $r_n \rightarrow 1$. Here, we are using the notation introduced in (4.1). Note that, for any $f \in A_{\beta}^p$, $\beta > -1$, we have

$$\int_0^1 M_p^p(f, r) r(1-r^2)^{\beta} dr < \infty,$$

and so

$$\begin{aligned} (1-r)^{\beta+1} M_p^p(f, r) &= (\beta+1) M_p^p(f, r) \int_r^1 (1-t)^{\beta} dt \\ &\leq (\beta+1) \int_r^1 M_p^p(f, t) (1-t)^{\beta} dt = o(1) \end{aligned}$$

as $r \rightarrow 1$. This, together with (6.3), yields $\varphi'_p \notin A_{p-2+c(2-p)^2}^p$. In other words, we have $\varphi_p \notin A_{p-2+c(2-p)^2, 1}^p$ by Proposition 2.2. Note that the hypothesis $p \geq 1$ is used here to assure that $p-2+c(2-p)^2 > -1$. Now, choose $s \in [0, 1)$ such that $sp-2+c(2-p)^2 > -1$ and put $\alpha = sp-2+c(2-p)^2$. Then, we have $s < \frac{\alpha+2}{p}$. Also, since $A_{p-2+c(2-p)^2, 1}^p \approx A_{\alpha, s}^p$ by (1.1), we have $\varphi_p \notin A_{\alpha, s}^p$. \square

The next example shows that the lower bound $s > \frac{\alpha+2}{p}$ in Theorem 3.3 is sharp when $\alpha > -1$.

Example 6.3. Let $p > 1$, $\alpha > -1$ and put $s = \frac{\alpha+2}{p}$. Then, there exists a holomorphic self-map φ of \mathbf{D} with $\varphi(1) = 1$ such that $C_{\varphi} : A_{\alpha, s}^p \rightarrow A_{\alpha, s}^p$ is bounded but φ does not have angular derivative at $z = 1$.

Proof. Let $\varphi = \varphi_{\eta}$ be any lens map. Note that $\alpha + (1-s)p > -1$. Thus, as in the proof of Proposition 5.4, we have

$$\mu_{p, \alpha+(1-s)p}(SI) \lesssim |I|^{p+(2+\alpha-sp)/\eta} = |I|^p,$$

so that $C_{\varphi} : A_{\alpha, s}^p \rightarrow A_{\alpha, s}^p$ is bounded by Theorem 2.6(i). Clearly, φ does not have an angular derivative at 1. \square

The next example shows that the upper bound for $\theta(P)$ in Theorem 5.5 is sharp when $\alpha_1 > -1$.

Example 6.4. Let p_j , s_j , α_j be as in the hypotheses of Theorem 5.5. Assume $\alpha_1 > -1$ and

$$(6.4) \quad \frac{p_1(\alpha_2 + 2 - s_2 p_2)}{p_2(\alpha_1 + 2 - s_1 p_1)} < \eta < 1.$$

Then, $f \circ \varphi_{\eta} \notin A_{\alpha_2, s_2}^{p_2}$ for some $f \in A_{\alpha_1, s_1}^{p_1}$.

Proof. Let $\varphi = \varphi_{\eta}$. Choose $0 < a < 1$ such that $|\varphi(a)| \geq 1/2$. Also, by using (6.4), choose $\varepsilon > 0$ sufficiently small so that

$$(6.5) \quad \frac{(2 + \alpha_2)/p_2 - s_2 + \varepsilon}{(2 + \alpha_1)/p_1 - s_1} < \eta < 1.$$

Now, consider the test function $f_a(z) = \log(1 - \overline{\varphi(a)}z)$. Let $k \geq s_1$ be a positive integer. Then we have $A_{\alpha_1, s_1}^{p_1} \approx A_{\alpha_1+(k-s_1)p, k}^{p_1}$ by (1.1). Therefore, by Proposition

2.2, (2.3), (6.2) and (6.5),

$$\begin{aligned}
 \|f_a\|_{A_{\alpha_1, s_1}^{p_1}} &\approx \|f_a\|_{A_{\alpha_1 + (k-s_1)p_1, k}^{p_1}} \\
 &\approx (1 - |\varphi(a)|^2)^{(2+\alpha_1)/p_1 - s_1} \\
 &\approx (1 - a)^{[(2+\alpha_1)/p_1 - s_1]\eta} \\
 (6.6) \quad &\lesssim (1 - a)^{[(2+\alpha_2)/p_2 - s_2] + \varepsilon}.
 \end{aligned}$$

On the other hand, for $\alpha_2 > -1$, we have by (1.1) and (3.2),

$$\begin{aligned}
 \|f_a \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}} &\approx \|f_a \circ \varphi\|_{A_{\alpha_2 + (1-s_2)p_2, 1}^{p_2}} \\
 &\gtrsim \frac{|\varphi'(a)|(1 - a)^{(2+\alpha_2)/p_2 + (1-s_2)}}{(1 - |\varphi(a)|^2)}
 \end{aligned}$$

and thus by (6.2),

$$(6.7) \quad \|f_a \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}} \gtrsim (1 - a)^{(2+\alpha_2)/p_2 - s_2}.$$

This also holds for $\alpha_2 = -1$, because $A_{-1, s_2}^{p_2} \subset A_{0, s_2}^{2p_2}$ by (1.4). Consequently, we obtain from (6.7), (6.6) and Lemma 4.3 that

$$\frac{\|f_a \circ \varphi\|_{A_{\alpha_2, s_2}^{p_2}}}{\|f_a\|_{A_{\alpha_1, s_1}^{p_1}}} \gtrsim (1 - a)^{-\varepsilon}.$$

Now, letting $a \rightarrow 1$, we see that C_φ does not take $A_{\alpha_1, s_1}^{p_1}$ into $A_{\alpha_2, s_2}^{p_2}$ by the Closed Graph Theorem. \square

REFERENCES

- [BB] F. Beatrous and J. Burbea, *Holomorphic Sobolev spaces on the ball*, Dissertationes Mathematicae, CCLXXVI (1989), 1-57. MR **90k**:32010
- [CM] C. Cowen and B. MacCluer, *Composition operators on spaces of analytic functions*, CRC Press, Boca Raton, FL, 1995. MR **97i**:47056
- [D] P. Duren, *Theory of H^p spaces*, Pure and Appl. Math., Vol. 38, Academic Press, New York, 1970. MR **42**:3552
- [G] J. Garnett, *Bounded analytic functions*, Pure and Appl. Math., vol. 96, Academic Press, New York, 1981. MR **83g**:30037
- [JM] P. Jones and N. Makarov, *Density properties of harmonic measure*, Annals of Mathematics, 142 (1995), 427-455. MR **96k**:30027
- [HKZ] H. Hedenmalm, B. Korenblum and K. Zhu, *Theory of Bergman spaces*, Graduate Texts in Math., vol. 199, Springer, New York, 2000. MR **2001c**:46043
- [L1] D. Luecking, *Forward and Reverse Carleson inequalities for functions in Bergman spaces and their derivatives*, Amer. J. Math., 107 (1985), 85-111. MR **86g**:30002
- [L2] D. Luecking, *Multipliers of Bergman spaces into Lebesgue spaces*, Proc. Edinburgh Math. Soc., 29 (1986), 125-131. MR **87e**:46034
- [MS] B. MacCluer and J. H. Shaprio, *Angular derivatives and compact composition operators on the Hardy and Bergman spaces*, Canad. J. Math., XXXVIII(4) (1986), MR **87h**:47048 878-906.
- [Ma] K. Madigan, *Composition operators on analytic Lipschitz spaces*, Proc. Amer. Math. Soc., 119 (1993), 465-473. MR **93k**:47043
- [Mi] J. Miao, *A property of analytic functions with Hadamard gaps*, Bull. Austral. Math. Soc., 45 (1992), 105-112. MR **93c**:30060
- [P] Ch. Pommerenke, *Boundary behavior of conformal maps*, Grundlehren der Mathematischen Wissenschaften, vol. 299, Springer-Verlag, Berlin, Heidelberg, New York, 1992. MR **95b**:30008
- [R1] W. Rudin, *Function theory in the unit ball of \mathbb{C}^n* , Grundlehren der Mathematischen Wissenschaften, vol. 241, Springer-Verlag, New York, 1980. MR **82i**:32002

- [R2] W. Rudin, *Real and complex analysis*, McGraw-Hill, New York, 1987. MR **88k**:00002
- [Sh] J.H. Shapiro, *Compact composition operators on spaces of boundary-regular holomorphic functions*, Proc. Amer. Math. Soc., 100(1) (1987), 49-57. MR **88c**:47059
- [Sm1] W. Smith, *Composition operators between Bergman and Hardy spaces*, Trans. Amer. Math. Soc., 348(6) (1996), 2331-2348. MR **96i**:47056
- [Sm2] W. Smith, *Compactness of composition operators on BMOA*, Proc. Amer. Math. Soc., 127(9) (1999), 2715-2725. MR **99m**:47040
- [SY] W. Smith and L. Yang, *Composition operators that improve integrability on weighted Bergman spaces*, Proc. Amer. Math. Soc., 126(2) (1998), 411-420. MR **98d**:47070
- [ST] J. Shapiro and P. Taylor, *Compact, nuclear, and Hilbert-Schmidt composition operators on H^2* , Indiana Univ. Math. J., 23(6) (1973), 471-496. MR **48**:4816

DEPARTMENT OF MATHEMATICS, KOREA UNIVERSITY, SEOUL 136-701, KOREA

E-mail address: `choebr@math.korea.ac.kr`

DEPARTMENT OF MATHEMATICS, KOREA UNIVERSITY, SEOUL 136-701, KOREA

E-mail address: `koohw@math.korea.ac.kr`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HAWAII, HONOLULU, HAWAII 96822

E-mail address: `wayne@math.hawaii.edu`

DISTRIBUTIONS OF CORANK 1 AND THEIR CHARACTERISTIC VECTOR FIELDS

B. JAKUBCZYK AND M. ZHITOMIRSKII

ABSTRACT. We prove that any 1-parameter family of corank 1 distributions (or Pfaff equations) on a compact manifold M^n is trivializable, i.e., transformable to a constant family by a family of diffeomorphisms, if all distributions of the family have the same characteristic line field. The characteristic line field is a field of tangent lines which is invariantly assigned to a corank one distribution. It is defined on M^n , if $n = 2k$, or on a subset of M^n called the Martinet hypersurface, if $n = 2k + 1$. Our second main result states that if two corank one distributions have the same characteristic line field and are close to each other, then they are equivalent via a diffeomorphism. This holds under a weak assumption on the singularities of the distributions. The second result implies that the abnormal curves of a distribution determine the equivalence class of the distribution, among distributions close to a given one.

0. INTRODUCTION

The well-known Gray theorem [G] states that any 1-parameter family of contact structures on a compact manifold M^{2k+1} is trivializable, i.e., transformable to a constant family by a family of diffeomorphisms. Our first main result generalizes this theorem to the case of “singular contact structures”, for which the contact condition is satisfied on a dense subset of M , and to corank one distributions on manifolds of even dimension. In these cases the family of distributions has to preserve, when the parameter changes, a characteristic line field. The *characteristic line field* is a field of tangent lines which is invariantly assigned to a corank one distribution (it is defined on M^n if $n = 2k$, or on a hypersurface of M^n if $n = 2k + 1$).

Our second main result states that if two corank one distributions have the same characteristic line field and are close to each other, then they are equivalent via a diffeomorphism. It means, in particular, that the characteristic line field contains complete information about the geometry of singularities of the distribution. Our results hold under a weak assumption on the distributions, called condition (A), saying that the depth of a characteristic ideal of the distribution is nondegenerate at singular points of the characteristic line field.

Received by the editors January 9, 2002 and, in revised form, September 4, 2002.

2000 *Mathematics Subject Classification.* Primary 58A17; Secondary 53B99.

Key words and phrases. Pfaff equation, equivalence, contact structure, quasi-contact structure, singularity, invariants, line field, homotopy method.

The first author was supported by the Committee for Scientific Research (KBN), Poland, grant 2P03A 03516.

The second author was supported by the Fund for the Promotion of Research at the Technion.

Let us state our second result in the simple case where the characteristic line field does not have singularities. Assume $n = 2k \geq 4$ and consider a smooth differential 1-form ω on M^n . Let $(\omega \wedge (d\omega)^{k-1})(p) \neq 0$ for all $p \in M^n$. Such a form ω defines the distribution $\Delta = \ker \omega$, called the *quasi-contact structure* defined by ω . Then $d\omega(p)$, restricted to the distribution $\Delta(p) = \ker \omega(p)$, is of maximal possible rank $2k - 2$ and has 1-dimensional kernel. We define the *characteristic line* at p by

$$L_p = \ker d\omega(p)|_{\Delta(p)}.$$

The *characteristic line field* $L = L(\omega)$ is the field of characteristic lines $p \rightarrow L_p$ on M . The following fact is a special case of our Theorem 1.2, the case where singularities are absent. (It is also a special case of a theorem in [MZh], Appendix A, concerning corank 1 distributions of constant class.)

Theorem 0.1. *Let Δ be the quasi-contact structure on a compact orientable manifold M^{2k} defined by a 1-form ω , and let $\tilde{\omega}$ be a 1-form such that $L(\tilde{\omega}) = L(\omega)$. If $\tilde{\omega}$ is sufficiently C^2 -close to ω , then there exists a diffeomorphism of M^{2k} sending $\tilde{\Delta} = \ker \tilde{\omega}$ to Δ .*

If a compact orientable manifold M^{2k} admits a quasi-contact structure defined by a global 1-form ω , then its characteristic line field is generated by a nonvanishing global vector field and the Euler characteristic of M^{2k} is equal to zero. Thus a manifold M^{2k} with nonzero Euler characteristic admits only corank one distributions with singular characteristic line field. Even if M^{2k} admits a quasi-contact structure, singularities may appear naturally when restricting a corank one distribution to a submanifold of even dimension. Therefore, it is natural to ask whether Theorem 0.1 holds in the presence of singularities. Theorem 1.2 in Section 1 gives a positive answer under assumption (A), saying that the singularities of the characteristic vector field have a natural depth (and codimension, in the analytic case). This assumption excludes singularities of infinite codimension.

To state a similar result in the case $n = 2k + 1$, we introduce the set of points where ω does not satisfy the contact condition:

$$S = \{p \in M^{2k+1} : (\omega \wedge (d\omega)^k)(p) = 0\}.$$

This set is called the *Martinet hypersurface*. The Martinet hypersurface is the set of zeros of the function

$$H = \omega \wedge (d\omega)^k / \Omega,$$

where Ω is a volume form. If S is empty, i.e., ω is a contact 1-form on M^{2k+1} , then $\Delta = \ker \omega$ is globally equivalent to any distribution $\tilde{\Delta}$ sufficiently close to Δ . This follows from the theorem of Gray mentioned above. Assume now that S is nonempty. We shall call $\Delta = \ker \omega$ a *Martinet distribution* if it satisfies the following two conditions:

- (a) $dH(p) \neq 0$ for all $p \in S$ (then S is smooth), and
- (b) $\Delta_S = \ker \omega_S$ is a quasi-contact structure on S , where ω_S is the pullback of ω to S .

At each point of S we can define the *characteristic line*

$$L_p = \ker d\omega_S(p)|_{\ker \omega_S(p)}, \quad p \in S.$$

The *characteristic line field* $L = L(\omega)$ for a Martinet distribution Δ is the field of tangent lines $p \rightarrow L_p$ on S . It has no singularities. The following fact is a special case of our Theorem 1.4.

Theorem 0.2. *Let $\Delta = \ker \omega$ be a Martinet distribution on a compact orientable manifold M^{2k+1} . If the Martinet hypersurfaces and the characteristic line fields of Δ and $\tilde{\Delta} = \ker \tilde{\omega}$ are the same and $\tilde{\omega}$ is sufficiently C^3 -close to ω , then there exists a diffeomorphism of the manifold sending $\tilde{\Delta}$ to Δ .*

Martinet distributions form a restrictive class of corank one distributions (L may have singularities). In particular, any Martinet distribution on M^3 has the Martinet hypersurface S , which consists of two-dimensional tori (each connected component of S has zero Euler characteristic, since S admits a 1-dimensional foliation defined by L). Our Theorem 1.4 will generalize Theorem 0.2 to the case of general corank one distributions and Pfaff equations on M^{2k+1} .

The appearance of the characteristic line field L as one of the main invariants of corank one distributions Δ has the following history. The first to study L was J. Martinet in [Mar] for the simplest occurring singularities of Δ on \mathbb{R}^{2n+1} (when L has no singularities). J. Martinet also started to study typical singularities of L in the 3-dimensional case. These singularities were roughly classified in [JP], where the existence of a modulus in the classification of characteristic line fields was shown. It was proved in [Zh1] by obtaining a normal form for Δ that this modulus is the only invariant of Δ . This gave the complete local classification of germs of generic 2-distributions on 3-manifolds, with the characteristic line field as the complete invariant.

In the book [Zh2] the second author gave a classification of finitely determined singularities of corank one distributions and Pfaff equations on manifolds of any dimension. In this case again the characteristic line field L is a complete invariant. This justifies, to a large extent, making the following

Conjecture. *In the space of germs at $0 \in \mathbb{R}^n$ of corank one distributions on \mathbb{R}^n , there is an exceptional set of infinite codimension such that for any two distributions Δ_0 and Δ_1 away from this set, with the same characteristic line field L , there exists a local diffeomorphism $\Phi : (\mathbb{R}^n, 0) \rightarrow (\mathbb{R}^n, 0)$ reducing Δ_1 to Δ_0 .*

In a weaker form, with $n = 3$, this conjecture has already appeared in [JP] and in a letter from J. Martinet to the second author (in 1989). In this case the conjecture can be deduced from the results in [JZh2], [JZh1]. The exceptional set consists of germs that do not satisfy the assumption (A) or that do not have the property of zeros (see Section 1). In [JZh2] we proved that away from the exceptional set the restriction of the distribution to the Martinet hypersurface is a complete invariant for any $n = 2k + 1$. If $k = 1$, then the restriction can be identified with the characteristic line field.

The results of the present paper concern global corank one distributions close to a fixed distribution. They also deal with families of distributions. In this setting we eliminate some of the difficulties in the above conjecture which are due to non-close germs and the necessity of preserving a fixed point (the source of the germ).

Our transition to the global approach was inspired by the results in [Gol] and [MZh, Appendix A]. In [Gol] it is proved that a family of Engel structures E_t on a 4-manifold is trivializable provided that the characteristic line field of E_t does not depend on t . The result in Appendix A of [MZh] states that two close global corank one distributions of any constant class (in the Cartan-Frobenius sense) are diffeomorphic provided that they have diffeomorphic characteristic foliations. This is a generalization of the Gray theorem. All these results apply to objects without

singularities, whereas in the present paper we allow any singularities except certain ones of infinite codimension, excluded by the assumption (A). The presence of singularities leads to the main difficulties in our proofs. In Section 2 we explain that the assumption (A) is natural and give examples showing that it cannot be weakened.

In [JZh1] and [JZh2] we obtained local realization theorems (for germs at a fixed point), theorems characterizing the set of all possible characteristic line fields if $n = 3$. Combined with the reduction theorems, they lead to a number of applications including classification results. In this paper we leave aside the difficult task of obtaining global realization theorems. The absence of such theorems restricts, at present, the possibility of drawing immediate interesting conclusions concerning global classification of corank one distributions.

Note that the assumption of closeness which appears in Theorems 0.1 and 0.2 (and later in Theorems 1.2 and 1.4) is essential for our results. Already two contact structures that are far from each other are, in general, not equivalent. A classification of contact structures is known only on certain 3-dimensional manifolds, see e.g. [TEG]. We hope that our results can be used in global contact or quasi-contact geometry for studying singularities of 1-forms which appear when two global contact (or quasi-contact) structures are joined by a path.

There are natural consequences of our results concerning characteristic curves of a distribution, also called *singular curves* or *abnormal curves* in sub-Riemannian geometry, and the geometry of distributions (cf. [A], [BH], [LS], [Mon]). These curves coincide in our case with the integral curves of the characteristic line field.

Under the assumptions of Theorems 0.1 and 0.2, as well as those of Theorems 1.2 and 1.4 in Section 1, the singular curves of a corank 1 distribution determine the equivalence class of the distribution, among distributions close to a given one in the C^∞ topology.

1. STATEMENT OF RESULTS

We will deal with Pfaff equations, which are more general objects than (cooriented) corank one distributions. Let M^n denote a compact, orientable, Hausdorff manifold of dimension $n \geq 3$. By definition, a Pfaff equation is a set of differential 1-forms on M^n generated, as a module over the ring of functions, by a single 1-form ω . In other words, a Pfaff equation is a 1-form on M^n defined up to multiplication by a nonvanishing function. We denote the Pfaff equation by $P = (\omega)$. If ω vanishes at no points of M^n , then (ω) can be identified with the field of kernels of ω —a coorientable hyperplane field in TM^n . In general, a Pfaff equation is a more general object, since we do not exclude the possibility of ω vanishing at some points of the manifold.

All objects in this paper will belong to a fixed category which is either C^∞ or real analytic C^ω .

The case $n = 2k$. To any Pfaff equation P , and in particular to any cooriented corank one distribution, one can associate the *characteristic line field*.

Definition 1.1. If n is even, $n = 2k$, then any vector field X defined by the relation

$$X \lrcorner \Omega = \omega \wedge (d\omega)^{k-1},$$

where Ω is any volume form, and ω any generator of P , is called a *characteristic vector field* of P . The line field on M^n generated by X , i.e., the mapping

$$p \rightarrow L_p = \{aX(p), a \in \mathbb{R}\} \subset T_p M^{2k},$$

is called the *characteristic line field*. The characteristic line field will be denoted by L or $L(\omega)$ or $L(P)$. The set $\{p \in M^{2k} : X(p) = 0\}$ is called the set of singular points of L and denoted by $Sing(L)$.

It is easy to check that any two characteristic vector fields differ by multiplication by a nonvanishing function, and consequently the characteristic line field is invariantly related to the Pfaff equation P , i.e., the choice of the generator ω of P and the volume form Ω is irrelevant. Note that if $(\omega \wedge (d\omega)^{k-1})(p) \neq 0$, then the definitions of L_p and L coincide with those given in the introduction.

In the presence of singularities we need an invariant that describes a "degree of degeneration" of singular points of a characteristic vector field X (for n odd it was introduced in [JZh2]). First we define an invariant that is slightly stronger than the set $Sing(L)$.

Let $p \in Sing(L)$. The *local characteristic ideal* I_p at a point $p \in M^{2k}$ of a Pfaff equation P is the ideal $I_p(X)$ in the ring of function germs at p , generated by the coefficients a_1, \dots, a_n of a characteristic vector field X of P , in some coordinate system near p . It is easy to see that the ideal I_p is invariantly related to the germ at p of $P = (\omega)$ (the choices of a characteristic vector field and of a local coordinate system are irrelevant). The germ at p of the set $Sing(L)$ is the zero set of I_p .

Definition 1.2. If $n = 2k$ and $p \in Sing(L)$, then we define $d_p(P) = d_p(X)$ as

$$d_p(P) = \text{depth } I_p(X).$$

Recall that the depth of a proper ideal $I \subset R$ of a ring R is the maximal length of a regular sequence of elements in I . A sequence $a_1, \dots, a_r \in I$ is called regular if a_1 is not a zero divisor in R and, for any $i = 2, \dots, r$, the element a_i is not a zero divisor in the quotient ring $R/(a_1, \dots, a_{i-1})$, where (a_1, \dots, a_{i-1}) denotes the ideal generated by a_1, \dots, a_{i-1} . By definition, $\text{depth } R = \infty$.

Remark. In the analytic category, $d_p(P)$ is equal to the codimension in \mathbb{C}^n of the germ at p of the set of complex zeros of the ideal I_p (i.e., the zero level set of the ideal generated by the complexification of the generators of I_p in some local coordinates). This follows from the fact that the complexification does not change the depth of an ideal of analytic function germs (cf. e.g. [E]) and from the equality of the $\text{depth}(I)$ and the codimension of the analytic set of zeros of I for any ideal I of holomorphic function germs.

We introduce the following crucial condition:

$$(A) \quad d_p(P) \geq 3 \text{ for any point } p \in Sing(L).$$

This condition is rather weak, in particular generic, as will be explained in Section 2.

The following theorems hold in the categories C^∞ and C^ω , with M a compact orientable manifold.

Theorem 1.1. Let P_t , $t \in [0, 1]$, be a family of Pfaff equations on M^{2k} , $k \geq 2$, that satisfies the following conditions.

(a) All P_t define the same characteristic line field $L = L(P_t)$.

(b) All P_t satisfy condition (A).

Then there exists a family Φ_t of diffeomorphisms of M^{2k} sending P_t to P_0 .

Theorem 1.2. Let $P_0 = (\omega_0)$ and $P_1 = (\omega_1)$ be Pfaff equations on M^{2k} , $k \geq 2$, that have the same characteristic line field $L = L(P_0) = L(P_1)$. Assume that condition (A) holds for P_0 . Then there exists a diffeomorphism Φ sending P_1 to P_0 provided that ω_1 is sufficiently close to ω_0 in the C^∞ topology.

In the above theorems as well as in Theorems 1.3 and 1.4 below, all objects are in the same category C^∞ or C^ω , including regularity with respect to the parameter t . The diffeomorphism in Theorem 1.2 can be taken C^∞ -close to the identity.

Remark (Closeness of ω_1 to ω_0). In many cases one can present a number $r < \infty$ (depending on P_0) such that the closeness in the C^∞ topology in Theorem 1.2 can be replaced by closeness in the C^r topology. See Theorem 0.1 and Theorem B.2 in Appendix B.

The case $n = 2k + 1$. The most basic invariant of a Pfaff equation $P = (\omega)$ on M^{2k+1} is the set

$$S = \{p \in M^{2k+1} : (\omega \wedge (d\omega)^k)(p) = 0\},$$

called the *Martinet hypersurface*, which consists of points at which ω is not a contact form. This set, invariantly related to P , is the zero level of the function

$$H = \omega \wedge (d\omega)^k / \Omega,$$

where Ω is a volume form.

The ideal (H) of the ring of functions on M^{2k+1} , generated by H , is also invariantly related to P . It is called the *Martinet ideal*.

The characteristic line field of $P = (\omega)$ on M^{2k+1} is defined on the set S .

Definition 1.3. Any vector field X on M^{2k+1} satisfying the relation

$$X \rfloor \Omega = \omega \wedge (d\omega)^{k-1} \wedge dH \mod (H),$$

where ω is any generator of P , and H is any generator of the Martinet ideal, will be called a *characteristic vector field* of P . The line field on $S = \{H = 0\}$ defined by the relation

$$p \rightarrow L_p = \{aX(p), a \in \mathbb{R}\}, \quad p \in S,$$

is called the *characteristic line field* of P . It will be denoted by L or $L(\omega)$ or $L(P)$. The set of singular points of L is defined as $Sing(L) = \{p \in M^{2k+1} : H(p) = 0, X(p) = 0\}$.

Above and in the rest of the paper the equality of two objects (vector fields, differential forms) mod (H) means that their difference is divisible over H in the space of objects of the same category. To check that the line field defined above is tangent to S , note that the definition of X implies that $X \rfloor dH = 0$ at any point of $S = \{H = 0\}$ and that X vanishes at any point of S at which the 1-form dH vanishes. Thus $X(p) \in T_p S$ at any point $p \in S$ at which S is a smooth hypersurface, and X vanishes at all other points. It is easy to check that the characteristic line field is invariantly related to P , i.e., the choices of the generator ω of P , the generator H of the Martinet ideal, and the volume form Ω are irrelevant.

Note that in the case of the Martinet singularity ($p \in S$ and $(\omega \wedge (d\omega)^{k-1} \wedge dH)(p) \neq 0$) the definitions of L_p and L coincide with those given in the introduction.

In order to deal with deeper singularities of P , namely those allowing singular points of L , we introduce our invariant $d_p(P)$ in the case of odd n as follows.

Let $p \in \text{Sing}(L)$. The *local characteristic ideal* I_p at a point $p \in M^{2k+1}$ of a Pfaff equation P is the ideal $I_p(H, X)$ in the ring of function germs at p generated by the germ H_p of a generator of the Martinet ideal and the coefficients a_1, \dots, a_n of a characteristic vector field X of P , in some coordinate system near p . The ideal I_p is invariantly related to the germ at p of $P = (\omega)$ (the choices of a characteristic vector field, a generator of the Martinet ideal, and of a local coordinate system are irrelevant). The germ at p of the set $\text{Sing}(L)$ is the zero set of I_p .

Definition 1.4. If $n = 2k + 1$ and $p \in \text{Sing}(L)$, then we define

$$d_p(P) = d_p(H, X)$$

as the maximal length of a regular sequence in the characteristic ideal $I_p(H, X)$, starting with the germ H_p as the first element.

Remarks. (a) In Noetherian rings all maximal regular sequences in I are of the same finite length. Moreover, any regular sequence can be completed to a maximal regular sequence. This implies that in the analytic category, independently of the parity of n , we have

$$d_p(P) = \text{depth}(I_p).$$

(b) Similarly to the case $n = 2k$, in the analytic category, $d_p(P)$ is equal to the codimension in \mathbb{C}^n of the germ at p of the set of complex zeros of the complexification of the ideal I_p .

To formulate our reduction theorem for the most general case, we need two properties of the Martinet ideal (H) : the property of zeros and the extension property.

Definition 1.5. The Martinet ideal (H) has the *property of zeros* if for any $p \in S = \{H = 0\}$ the ideal in the ring of all function germs at p generated by the germ H_p of H at p coincides with the ideal in the same ring consisting of function germs vanishing on the germ at p of the set $S = \{H = 0\}$.

The property of zeros allows us to identify the Martinet hypersurface $S = \{H = 0\}$ with the Martinet ideal. In the case of germs this follows from the definition. Examples where the property of zeros is violated at a point p include:

- (a) $H_p = H_1^2 H_2$, where H_1, H_2 are function germs and $H_1(p) = 0$;
- (b) H_p is equivalent to $r^2 = x_1^2 + \dots + x_n^2$;
- (c) H_p is a flat germ (i.e., the Taylor series of H at p is zero);
- (d) H_p is a zero divisor in the ring of all germs at p .

In case (c) the property of zeros is violated, since $\tilde{H}_p = r^{-2} H_p$ is smooth and has the same germ of zeros as H_p but $\tilde{H}_p \notin (H_p)$. Note that (d) is a particular case of (c).

The local version of the property of zeros (Definition 1.5) implies the global version: if a function f on M vanishes on the set $S = \{H = 0\}$, then f belongs to the ideal (H) . This follows from the fact that division by H or by the germ H_p is unique (by (d) the germ H_p is not a zero divisor).¹

¹In the C^∞ -category the global and local versions of the property of zeros are equivalent (this follows from the partition of unity). In the real analytic category they are equivalent provided that the sheaf of functions vanishing on $S = \{H = 0\}$ is coherent. In the proof "global implies

We also need the *extension property* of the Martinet hypersurface $S = \{H = 0\}$. Denote by $C^\infty(M)$ the Fréchet space of smooth functions on M , equipped with the topology of convergence together with all derivatives. Let $C^\infty(M, S)$ denote its closed subspace of functions that vanish on S . We define the space of smooth functions on S as the quotient Fréchet space $C^\infty(S) = C^\infty(M)/C^\infty(M, S)$.

Definition 1.6. We say that S has the *extension property* if there exists a continuous linear operator $\lambda : C^\infty(S) \rightarrow C^\infty(M)$ such that $\lambda(f)|_S = f$ for all $f \in C^\infty(S)$.

The extension property automatically holds in the C^ω category, since it holds for any analytic subset S of M (see [BS] for a more general extension theorem). It also holds if we assume that (H) has, locally around any point $p \in S$, a generator that is analytic in some coordinate system.

The following theorems hold in the categories C^∞ and C^ω , with M compact and orientable.

Theorem 1.3. Let P_t , $t \in [0, 1]$, be a family of Pfaff equations on M^{2k+1} , $k \geq 1$, that satisfies the following conditions.

(a) All P_t have the same Martinet hypersurface S , which has the extension property, and their Martinet ideals have the property of zeros (and consequently are the same).

(b) All P_t define the same characteristic line field $L = L(P_t)$.

(c) All P_t satisfy condition (A).

Then there exists a family Φ_t of diffeomorphisms of M^{2k+1} sending P_t to P_0 .

Remark. Recall that the extension property of S holds automatically in the C^ω category. We conjecture that in the C^∞ category the extension property in Theorem 1.3 also can be omitted. Our proofs show that this is so if the family P_t has a generator ω_t that is polynomial in t .

Theorem 1.4. Let $P_0 = (\omega_0)$ and $P_1 = (\omega_1)$ be Pfaff equations on M^{2k+1} , $k \geq 1$, which have the same Martinet hypersurface $S = S(P_0) = S(P_1)$ and the same characteristic line field $L = L(P_0) = L(P_1)$. Assume that the Martinet ideal of P_0 has the property of zeros and P_0 satisfies condition (A). Then there exists a diffeomorphism Φ sending P_1 to P_0 , provided that ω_1 is sufficiently close to ω_0 in the C^∞ topology.

Remark (Closeness of ω_1 to ω_0). As in the even-dimensional case, often one can present a number $r < \infty$ (depending on P_0) such that the closeness in the C^∞ topology in Theorem 1.4 can be replaced by closeness in the C^r topology. See Theorem 0.2 and Theorem B.2 in Appendix B.

The contents of the further sections. In Section 2 we explain why the condition (A) is natural and give examples showing that it cannot be weakened. The consequences of condition (A) are explained in Section 3 and Appendix A: the condition (A) implies certain global division properties of a characteristic vector field. Section 4 contains auxiliary algebraic statements, which also will be used throughout the proofs. Using the division properties and these algebraic statements, we prove Theorems 1.1 and 1.3 in Sections 5 and 7, respectively. The proofs of these theorems are based on the homotopy method, according to which it suffices to prove

local" one should use Cartan's Theorem A in [C], which says that any local section of a coherent analytic module belongs to the module generated by global sections.

the solvability of the equation

$$(HE) \quad L_{Z_t}\omega_t + h_t\omega_t + \frac{d\omega_t}{dt} = 0$$

with respect to a family of vector fields Z_t and a family of functions h_t (here $L_Z\omega$ denotes the Lie derivative of ω along Z). Then the family Φ_t of diffeomorphisms obtained by integrating the family of vector fields Z_t ,

$$\frac{d\Phi_t}{dt} = Z_t(\Phi_t), \quad \Phi_0 = id,$$

transforms the Pfaff equations (ω_t) into (ω_0) :

$$\Phi_t^*\omega_t = \psi_t\omega_0,$$

where $\psi_t = \exp(-\int_0^t \tilde{h}_s ds)$ and $\tilde{h}_t = h_t \circ \Phi_t$. In what follows the equation (HE) will be called the *homotopy equation* or *homological equation*.

Theorems 1.2 and 1.4 are proved in Sections 6 and 8 by reduction to Theorems 1.1 and 1.3. In these sections we show that if Pfaff equations P_0 and P_1 satisfy the assumptions of Theorem 1.2 or 1.4, then there exist generators ω_0 of P_0 and ω_1 of P_1 such that the path of Pfaff equations P_t generated by $\omega_t = \omega_0 + t(\omega_1 - \omega_0)$ satisfies the assumptions of Theorem 1.1 or Theorem 1.3.

In Appendix B we present certain topological properties of linear operators related to the Martinet ideal and the characteristic ideal. We also show a way of transition from the assumption of C^∞ -closeness of ω_1 to ω_0 in Theorems 1.2 and 1.4 to the C^r -closeness with a certain $r < \infty$. In the simplest cases this way leads to Theorems 0.1 and 0.2 in the Introduction.

2. NECESSITY OF CONDITION (A)

In this section we explain why the condition

$$(A) \quad d_p(P) \geq 3$$

is natural, and we give examples showing that this condition cannot be weakened: if $\text{depth } d_p(P) = 2$, then our theorems are not true anymore.

Fix a point $p \in M^n$ and denote by J_p^i the space of i -jets of 1-forms at p . The condition that p is a singular point of the characteristic line field L , i.e. $p \in \text{Sing}(L)$, is the condition

$$(\omega \wedge (d\omega)^{k-1})(p) = 0,$$

if $n = 2k$, and

$$(\omega \wedge (d\omega)^k)(p) = 0, \quad (\omega \wedge (d\omega)^{k-1} \wedge dH)(p) = 0,$$

if $n = 2k + 1$. It involves the i -jet at p of a generator ω of P , where $i = 1$ if n is even and $i = 2$ if n is odd. This condition distinguishes a certain subset of J_p^i —the space of i -jets of ω at p . It is not difficult to see that for any parity of n this subset is a stratified submanifold of codimension 3 (see [Mar], [Zh2], or [JZh2] for more details). Consequently, for generic ω the set $\text{Sing}(L)$ is either empty or a submanifold of M^n of codimension 3. In the real analytic category (and conjecturally, in the smooth category too) the set of 1-forms ω violating (A) has infinite codimension in the space of all 1-forms on M^n ; see [JZh2], Proposition 3.4 and Theorem A2 (Appendix 2).

The following examples show that the condition (A) cannot be replaced by the condition $d_p(P) \geq 2$. In these examples $\dim M = 4$ and $\dim M = 5$. They can be

extended to higher dimensions. We have not found an example in the 3-dimensional case, but we believe that such an example exists.

Example 1. Consider the family of Pfaff equations on the 4-torus T^4 generated by 1-forms

$$\omega_t = d\theta_1 + (\sin \theta_3 \sin \theta_4 + t)d\theta_2.$$

The characteristic vector field X_t is the same for all t :

$$X_t = X_0 = \cos \theta_4 \sin \theta_3 \frac{\partial}{\partial \theta_3} - \cos \theta_3 \sin \theta_4 \frac{\partial}{\partial \theta_4}.$$

The set S of singular points of X_0 is the union of 8 disjoint 2-dimensional tori (4 of them are described by the equations $\theta_3, \theta_4 \in \{\pi/2, -\pi/2\}$, and the other 4 by the equations $\theta_3, \theta_4 \in \{0, -\pi\}$). The codimension of S is 2; therefore $\text{depth } I_p = 2$. The restriction of (ω_t) to any of these 2-dimensional tori is a Pfaff equation generated by a 1-form $\alpha_t = d\theta_1 + (\delta + t)d\theta_2$, where $\delta \in \{0, \pm 1\}$ depending on the torus. This Pfaff equation can be identified with the vector field $V_t : (\delta + t)\partial/\partial \theta_1 - \partial/\partial \theta_2$ defined up to multiplication by a nonvanishing function. It follows that the phase portrait of V_t on the torus is invariantly related to (ω_t) . It is well known that the equivalence of the phase portraits of V_{t_1} and V_{t_2} with a fixed δ implies $t_1 = t_2$ provided that t_2 is close to t_1 ; see [ArII] (the parameter t corresponds to the rotation number). Therefore the parameter t of the family P_t is a modulus (a parameter varying continuously and distinguishing nonequivalent Pfaff equations).

Example 2. Consider the family of Pfaff equations on the 5-torus

$$T^5(\theta_1, \theta_2, \phi_1, \phi_2, \phi_3)$$

generated by 1-forms

$$\omega_t = (A(\theta_1, \theta_2) + B_t(\theta_1, \theta_2, \phi_2))d\phi_1 + C(\theta_1, \theta_2)d\phi_2 + d\phi_3,$$

where

$$A(\theta_1, \theta_2) = 3(\sin \theta_1 + \sin \theta_2),$$

$$B_t(\theta_1, \theta_2) = t \sin \phi_2 (1 - \cos(\theta_1 - \theta_2)),$$

$$C(\theta_1, \theta_2) = \cos \theta_1 + \cos \theta_2.$$

A simple calculation gives that $\omega_t \wedge (d\omega_t)^2 = \sin(\theta_1 - \theta_2) \cdot Q_t \cdot \Omega$, where Ω is a volume form and Q_t is a family of nonvanishing functions on T^5 , if $t \in [-1, 1]$. Therefore the Martinet ideal is the same for all t ; it is generated by the function $\sin(\theta_1 - \theta_2)$. The Martinet hypersurface consists of two disjoint 4-tori:

$$S = T_1^4 \cup T_2^4, \quad T_1^4 = \{\theta_2 = \theta_1\}, \quad T_2^4 = \{\theta_2 = \theta_1 + \pi\}.$$

Since the function B_t vanishes on the torus T_1^4 , the restriction of (ω_t) to T_1^4 does not depend on t . The restriction of (ω_t) to the torus $T_2^4(\phi_1, \phi_2, \phi_3, \theta_1)$ depends on t ; it is the Pfaff equation (α_t) , where $(\alpha_t) = 2t \sin \phi_2 d\phi_1 + d\phi_3$. The characteristic vector field of (ω_t) restricted to T_2^4 is the characteristic vector field of (α_t) . It is $2t \cos \phi_2 \partial/\partial \theta_1$. Assume that $t \neq 0$. Then the characteristic line field does not depend on t . The set of its singular points is the union of two disjoint 3-tori T_{\pm}^3 , given by the equations $\phi_2 = \pm \pi/2$. The restriction of (ω_t) to T_{\pm}^3 (or, the same, the restriction of (α_t) to T_{\pm}^3) is the Pfaff equation of the form (β_t) , $\beta_t = 2t\delta d\phi_1 + d\phi_3$, $\delta \in \{-1, 1\}$. Consider the vector field $V_t : -\partial/\partial \phi_1 + 2t\delta \partial/\partial \phi_3$ on the 2-torus $T^2 = T^2(\phi_1, \phi_3)$. It is easy to see that the Pfaff equations (β_{t_1}) and (β_{t_2}) on the

3-torus are equivalent if and only if the phase portraits of V_{t_1} and V_{t_2} on the 2-torus are equivalent. As in the previous example, this is so if and only if $t_1 = t_2$ provided that t_2 is close to t_1 . Therefore the parameter t of the family (ω_t) of Pfaff equations is a modulus, although these Pfaff equations have the same Martinet ideal (satisfying the property of zeros) and the same characteristic line field. The reason for that is the violation of the assumption (A)—the depth of the characteristic ideal is equal to 2 instead of 3.

3. CONDITION (A) AND DIVISION PROPERTIES

In this section we explain implications of condition (A) which will be essential in further proofs. The main implications are the following global division properties of a characteristic vector field X . As before, we work in the C^∞ and C^ω categories.

Proposition 3.1.a. *If a Pfaff equation $P = (\omega)$ satisfies condition (A), then any characteristic vector field X of P has the following division properties.*

(i) *If n is even, then for any vector field Y and any r -form ν on M^n with $r = n - 1$ or $r = n - 2$ the equality*

$$X \lrcorner \nu = 0 \quad \text{implies} \quad \nu = X \lrcorner \mu,$$

for an $(r + 1)$ -form μ on M^n , and the equality

$$X \wedge Y = 0 \quad \text{implies} \quad Y = fX$$

for a function f on M^n .

(ii) *If n is odd, then for any vector field Y on M^n and any $(n - 1)$ -form ν on M^n the equality*

$$X \lrcorner \nu = 0 \mod (H) \quad \text{implies} \quad \nu = X \lrcorner \mu \mod (H),$$

for an n -form μ on M^n , and the equality

$$X \wedge Y = 0 \mod (H) \quad \text{implies} \quad Y = fX \mod (H)$$

for a function f on M^n . Here (H) is the Martinet ideal of P , and we assume that (H) has the property of zeros.

This proposition is a corollary of a general theorem in [DJ] on division properties of the interior product with a section X of a vector bundle (see Appendix A for the proof).

We also need a division property with parameters. In the next and all further statements in this section a 1-parameter family of functions, differential forms or vector fields on M is assumed to be *regular in t* , i.e., depending on t analytically (in the C^ω category) or smoothly (in the C^∞ category).

Proposition 3.1.b. *Proposition 3.1.a holds with the forms ν, μ , vector field Y , and function f replaced by families ν_t, μ_t, Y_t, f_t , $t \in [0, 1]$, provided that in the odd-dimensional case either the set $S = \{H = 0\}$ has the extension property (see Section 1) or the families ν_t and Y_t depend on t polynomially.*

This proposition is also proved in Appendix A, using the already mentioned general theorem on division properties.

Remark. Proposition 3.1 also holds for germs at a fixed point.

Another implication of condition (A) concerns the structure of the set $Sing(L)$ of singular points of the characteristic foliation L : it cannot be too degenerate.

Proposition 3.2. *If a Pfaff equation (ω) satisfies condition (A) and in the odd-dimensional case the Martinet ideal of (ω) has the property of zeros, then any characteristic vector field X of (ω) , the Martinet hypersurface S and the set $\text{Sing}(L)$ of singular points of the characteristic foliation have the following properties.*

(i) *If $n = 2k$, then the set $M^n \setminus \text{Sing}(L)$ of points where X does not vanish (i.e. (ω) is quasi-contact) is dense in M^n .*

(ii) *If $n = 2k + 1$, then the set $M^n \setminus S$ (i.e., the set of points at which ω is contact) is dense in M^n . Equivalently, any generator H of the Martinet ideal is not a zero divisor.*

(iii) *If $n = 2k + 1$, then the set $S \setminus \text{Sing}(L)$ is dense in S .*

Proof. Statement (i) follows from the observation that if ω is not quasi-contact at any point of an open set, then any characteristic vector field X vanishes on this set (vanishes on a connected component of M^n , in the analytic category). Consequently, given a point p in this set, the characteristic ideal I_p at p generated by the coefficients of X is trivial and contains no non-zero-divisor. This contradicts assumption (A).

Statement (ii) is a simple implication of the property of zeros of the Martinet ideal; see Definition 1.5 and the examples following it.

To prove (iii), assume that there exists a neighbourhood U in M^n of a point $p \in S$ such that a characteristic vector field X vanishes at any point of the set $U \cap S$. By the property of zeros of the Martinet ideal we obtain that $X_p = 0 \bmod(H_p)$, where the subscript indicates the germ at p . This contradicts assumption (A) at the point p . The proof is complete. \square

Propositions 3.1 and 3.2 imply the possibility of choosing the same characteristic vector field for all Pfaff equations with the same characteristic foliation.

Proposition 3.3. *Let $P_t = (\omega_t)$, $t \in [0, 1]$, be a family of Pfaff equations on M^n satisfying the assumptions of Theorem 1.1, if $n = 2k$, or of Theorem 1.3, if $n = 2k + 1$. Then for any family X_t of characteristic vector fields of (ω_t) we have*

$$X_t = R_t X_0, \quad \text{if } n = 2k, \quad \text{or}$$

$$X_t = R_t X_0 \bmod(H), \quad \text{if } n = 2k + 1,$$

where R_t , $t \in [0, 1]$, is a family of positive-valued functions and X_0 is X_t with $t = 0$.

Proof. Let $n = 2k$. The equality $L(\omega_t) = L(\omega_0)$ implies that $(X_t \wedge X_0)(p) = 0$ for all $p \in M^n$. By Proposition 3.1.b we obtain that $X_t = R_t X_0$, where R_t is a family of functions. Proposition 3.1.b also implies that for any fixed t we have $X_0 = Q_t X_t$, where Q_t is a function on M^n . This leads to the relation $(1 - R_t Q_t)X_0 = 0$. By Proposition 3.2, X_0 does not vanish on a dense subset of M^n ; thus $R_t Q_t = 1$. This implies that R_t is a family of nonvanishing functions. This family is positive valued, since for any $p \in M^n$ the function $R_t(p)$ is continuous in t and $R_0(p) = 1$.

In the case of $n = 2k + 1$ the equality $L(\omega_t) = L(\omega_0)$ gives $(X_t \wedge X_0)(p) = 0$ for all $p \in S$. From the property of zeros of the Martinet ideal we deduce that $X_t \wedge X_0 = 0 \bmod(H)$. Using Proposition 3.1.b, we see that $X_t = R_t X_0 \bmod(H)$, for a family of functions R_t . Similarly, we have $X_0 = Q_t X_t \bmod(H)$ for any fixed t , where the Q_t are functions. Therefore, $(1 - R_t Q_t)X_0 = 0 \bmod(H)$. By Proposition 3.2, $R_t Q_t = 1$ on S , and so R_t is nonvanishing on S . From the fact that $R_t(p)$ is continuous in t and from $R_0(p) = 1$, we deduce that R_t is positive valued on S .

for any $t \in [0, 1]$. Finally, adding to R_t the function CH^2 with a sufficiently large constant C , we obtain R_t positive valued on M . \square

Condition (A) implies one more division property that we need in our proofs. Its proof is postponed to Section 4.

Proposition 3.4. *Let $P_t = (\omega_t)$, $t \in [0, 1]$, be a family of Pfaff equations on M^{2k} satisfying the assumptions of Theorem 1.1, and let β_t be a family of 1-forms such that $\omega_t \wedge \beta_t = 0$. Then $\beta_t = h_t \omega_t$ for some family h_t of functions.*

Note that this statement is trivial if ω_t is a family of nonvanishing 1-forms, but we do not assume this in our theorems.

4. AUXILIARY ALGEBRAIC LEMMAS

To prove the solvability of the homotopy equation in the proofs of Theorems 1.1 and 1.3, we will also use the following simple algebraic facts.

Recall that a 1-form α on M^n , $n \geq 3$, is called contact (quasi-contact) at $p \in M^n$ if $n = 2k + 1$ (respectively, $n = 2k$) and $(\alpha \wedge (d\alpha)^k)(p) \neq 0$ (respectively, $(\alpha \wedge (d\alpha)^{k-1})(p) \neq 0$).

Lemma 4.1. *Let α and λ be 1-forms on M^{2k} . If α is quasi-contact at p and $(\lambda \wedge \alpha \wedge (d\alpha)^{k-2})(p) = 0$, then $(\lambda \wedge \alpha)(p) = 0$.*

Lemma 4.2. *Let α be a 1-form on M^{2k+1} . If α is a contact at p and λ is a 1-form such that $(\lambda \wedge \alpha \wedge (d\alpha)^{k-1})(p) = 0$ and $(\lambda \wedge (d\alpha)^k)(p) = 0$, then $\lambda(p) = 0$.*

The facts stated in these lemmas are invariant with respect to multiplication of α by a nonvanishing function, i.e., they are properties of the Pfaff equation (α) . These properties can be easily checked in the Darboux coordinates in which the Pfaff equation takes the form $(dz + x_1 dy_1 + \cdots + x_r dy_r)$, where $r = k - 1$ if $n = 2k$, and $r = k$ if $n = 2k + 1$.

Lemma 4.3. *Let α be a 1-form on M^{2k+1} that is not contact at p , but $\alpha(p) \neq 0$. If λ is a 1-form such that $(\lambda \wedge \alpha \wedge (d\alpha)^{k-1})(p) = 0$, then $(\lambda \wedge (d\alpha)^k)(p) = 0$.*

Proof. We take a nonzero vector $v \in T_p M^{2k+1}$ such that $v \lrcorner \alpha = v \lrcorner d\alpha = 0$. (The existence of such a vector follows from the assumption that α is not contact at p .) Then the relation assumed in the lemma implies that the form $(v \lrcorner \lambda) \cdot \alpha \wedge (d\alpha)^{k-1}$ vanishes at p . It follows that if $(\alpha \wedge (d\alpha)^{k-1})(p) \neq 0$, then $(v \lrcorner \lambda)(p) = 0$, and consequently $(\lambda \wedge (d\alpha)^k)(p) = 0$. On the other hand, if $(\alpha \wedge (d\alpha)^{k-1})(p) = 0$, then the assumption $\alpha(p) \neq 0$ implies that $(d\alpha)^k(p) = 0$, and then again $(\lambda \wedge (d\alpha)^k)(p) = 0$. \square

Lemma 4.4. *If Ω is a volume form on M^n , λ is a 1-form, γ is an $(n - 2)$ -form and X is a vector field defined by the relation $X \lrcorner \Omega = \lambda \wedge \gamma$, then $X \lrcorner \lambda = 0$.*

Proof. To prove this statement, note that the definition of X implies $X \lrcorner (\lambda \wedge \gamma) = 0$, and consequently $(X \lrcorner \lambda) \cdot \gamma \pm (X \lrcorner \gamma) \wedge \lambda = 0$. It follows that $(X \lrcorner \lambda) \cdot (\lambda \wedge \gamma) = 0$. Since X vanishes exactly at points at which the form $\lambda \wedge \gamma$ vanishes, we obtain that $X \lrcorner \lambda = 0$. \square

Finally, we need the following general properties of a characteristic vector field.

Lemma 4.5. *If X is a characteristic vector field of a Pfaff equation (ω) on M^{2k} , then*

$$X \rfloor \omega = 0 \quad \text{and} \quad (X \rfloor d\omega) \wedge \omega = 0.$$

Proof. The first relation follows from the definition of X and Lemma 4.4. The definition of X implies that $X \rfloor (\omega \wedge (d\omega)^{k-1}) = 0$, which, together with $X \rfloor \omega = 0$, gives $(X \rfloor d\omega) \wedge \omega \wedge (d\omega)^{k-2} = 0$ if $k > 1$. Now the second relation in Lemma 4.5 follows from Lemma 4.1 at points where ω is quasi-contact. At all other points the field X vanishes, and there is nothing to prove. \square

Lemma 4.6. *If X is a characteristic vector field of a Pfaff equation (ω) on M^{2k+1} whose Martinet ideal has the property of zeros, and H is a generator of this ideal, then*

$$X \rfloor \omega = 0 \pmod{(H)}, \quad X \rfloor dH = 0 \pmod{(H)}, \quad (X \rfloor d\omega) \wedge \omega = 0 \pmod{(H)}.$$

Proof. Due to the property of zeros of (H) , it suffices to prove the three relations at any point p of the Martinet hypersurface S such that $X(p) \neq 0$. The relation $(X \rfloor dH)(p) = 0$ follows immediately from the definition of X . To see the other two relations, note that S is regular in a neighbourhood of a point p such that $X(p) \neq 0$. The definition of the characteristic vector field X in the case $n = 2k + 1$ implies that the vector field $X|_S$ on S is, in a neighbourhood of such a point p , a characteristic vector field of the Pfaff equation $(\omega|_S)$ on S (which is quasi-contact at p). Thus the remaining two relations follow from Lemma 4.5. The proof is complete. \square

Proof of Proposition 3.4. Let X_t be the characteristic vector field of (ω_t) defined by $X_t \rfloor \Omega = \omega_t \wedge (d\omega_t)^{k-1}$. Since X_t (and so ω_t) does not vanish on a dense subset of M^n , the condition $\omega_t \wedge \beta_t = 0$ and Lemma 4.5 imply that $X_t \rfloor \beta_t = 0$ and $X_t \rfloor (\beta_t \wedge (d\omega_t)^{k-1}) = 0$. From Proposition 3.3 we have the equality $X_t = R_t X_0$, with R_t nonvanishing; thus $X_0 \rfloor (\beta_t \wedge (d\omega_t)^{k-1}) = 0$. Therefore the division property in Proposition 3.1.b implies the following relation: $\beta_t \wedge (d\omega_t)^{k-1} = X_0 \rfloor \mu_t = (g_t/R_t) X_t \rfloor \Omega$, where Ω is a volume form, g_t is a family of functions and $\mu_t = g_t \Omega$. Taking $h_t = g_t/R_t$, we can rewrite this relation in the form

$$(\beta_t - h_t \omega_t) \wedge (d\omega_t)^{k-1} = 0.$$

Let us show that this relation implies $\beta_t - h_t \omega_t = 0$. We know that $(\beta_t - h_t \omega_t) \wedge \omega_t = 0$, since $\beta_t \wedge \omega_t = 0$. Fix t and a point p at which ω_t is quasi-contact. At this point ω_t does not vanish; therefore $(\beta_t - h_t \omega_t)(p) = r \omega_t(p)$, with the scalar r depending on t and p . Then the displayed relation implies that $r(\omega_t \wedge (d\omega_t)^{k-1})(p) = 0$ and consequently $r = 0$. So, $(\beta_t - h_t \omega_t)(p) = 0$ if p is a point at which ω_t is quasi-contact. By Proposition 3.2, (i) the set of such points is dense, and so $\beta_t = h_t \omega_t$ at any point of the manifold. The proof is complete. \square

Now we are ready to prove the solvability of the homotopy equation (HE) and our main theorems.

5. PROOF OF THEOREM 1.1

Solvability of the homotopy equation (HE) in Section 1 is equivalent to solvability of

$$(5.1) \quad \left(L_{Z_t} \omega_t + \frac{d\omega_t}{dt} \right) \wedge \omega_t \wedge (d\omega_t)^{k-2} = 0,$$

with respect to a family Z_t of vector fields. Namely, equation (5.1) is obtained from the homotopy equation by external multiplication by $\omega_t \wedge (d\omega_t)^{k-2}$. Conversely, if (5.1) is solvable then, using the fact that the set of quasi-contact points of (ω_t) is dense in M^n (Proposition 3.2, (i)), we get from (5.1) by Lemma 4.1 that $(L_{Z_t}\omega_t + (d\omega_t/dt)) \wedge \omega_t = 0$. Therefore, by Proposition 3.4 we get $L_{Z_t}\omega_t + (d\omega_t/dt) + h_t\omega_t = 0$, for a family of functions h_t , which is the homotopy equation (HE).

A solution Z_t of equation (5.1) will be constructed within the set of families Z_t satisfying

$$(5.2) \quad Z_t \rfloor \omega_t = 0.$$

Condition (5.2) implies that $L_{Z_t}\omega_t = Z_t \rfloor d\omega_t$, and the equation (5.1) can be rewritten in the form

$$(5.3) \quad Z_t \rfloor (\omega_t \wedge (d\omega_t)^{k-1}) + (k-1) \frac{d\omega_t}{dt} \wedge \omega_t \wedge (d\omega_t)^{k-2} = 0.$$

In order to solve equation (5.3) we fix a volume form Ω and define a family X_t of characteristic vector fields of (ω_t) by the relation $X_t \rfloor \Omega = \omega_t \wedge (d\omega_t)^{k-1}$. Lemma 4.5 and Proposition 3.3 imply the relations $X_0 \rfloor \omega_t = 0$, $X_0 \rfloor (d\omega_t/dt) = 0$ and $(X_0 \rfloor d\omega_t) \wedge \omega_t = 0$. Thus

$$X_0 \rfloor \nu_t = 0, \quad \text{where} \quad \nu_t = \left(\frac{d\omega_t}{dt} \wedge \omega_t \wedge (d\omega_t)^{k-2} \right).$$

Therefore, by the division property in Proposition 3.1.b, we have

$$(5.4) \quad \nu_t = X_0 \rfloor \tilde{\mu}_t,$$

with some family $\tilde{\mu}_t$ of $(n-1)$ -forms of the same regularity with respect to t as in ω_t . Using Proposition 3.3 again, we obtain

$$(5.5) \quad \nu_t = X_t \rfloor \mu_t$$

for some, regular in t , family of $(n-1)$ -forms μ_t . This relation allows to rewrite equation (5.3) in the form

$$Z_t \rfloor (X_t \rfloor \Omega) + (k-1)X_t \rfloor \mu_t = 0.$$

The latter equation has a solution Z_t defined by the relation

$$Z_t \rfloor \Omega = (k-1)\mu_t.$$

It now remains to check that the constructed solution Z_t satisfies relation (5.2). The equality (5.2) is equivalent to the relation $\mu_t \wedge \omega_t = 0$. From (5.5) and the definition of ν_t we have $(X_t \rfloor \mu_t) \wedge \omega_t = 0$. By Lemma 4.5, $X_t \rfloor \omega_t = 0$; therefore $X_t \rfloor (\mu_t \wedge \omega_t) = 0$. Thus the n -form $\mu_t \wedge \omega_t$ vanishes at any point at which X_t does not vanish. By Proposition 3.2, (i) the set of such points is everywhere dense; therefore $\mu_t \wedge \omega_t = 0$ and (5.2) holds. This completes the proof of Theorem 1.1.

6. PROOF OF THEOREM 1.2

We will use the following proposition (its proof is postponed to the end of this section).

Proposition 6.1. *Assume that $P_0 = (\omega_0)$ satisfies condition (A) and*

$$(6.1) \quad \omega_1 \wedge (d\omega_1)^{k-1} = \omega_0 \wedge (d\omega_0)^{k-1}.$$

Then for the path $\omega_t = (1-t)\omega_0 + t\omega_1$ we have

$$(6.2) \quad \omega_t \wedge (d\omega_t)^{k-1} = A_t \omega_0 \wedge (d\omega_0)^{k-1},$$

where A_t is a family of functions, polynomial in t .

Proof of Theorem 1.2. The equality $L(\omega_0) = L(\omega_1)$ implies $X_0 \wedge X_1 = 0$, where X_0 and X_1 are characteristic vector fields of P_0 and P_1 . Condition (A) satisfied for (ω_0) enables us to use the second division property in Proposition 3.1.a, (i) to deduce that $X_1 = RX_0$ and, equivalently,

$$(6.3) \quad \omega_1 \wedge (d\omega_1)^{k-1} = R\omega_0 \wedge (d\omega_0)^{k-1},$$

where R is a smooth or analytic function. In fact, R is positive valued, which will follow from the closeness of ω_1 to ω_0 . Therefore, assuming $R > 0$, we choose the generator

$$\hat{\omega}_1 = R^{1/k}\omega_1,$$

and we have

$$(6.4) \quad \hat{\omega}_1 \wedge (d\hat{\omega}_1)^{k-1} = \omega_0 \wedge (d\omega_0)^{k-1}.$$

Let

$$(6.5) \quad \omega_t = (1-t)\omega_0 + t\hat{\omega}_1.$$

To prove Theorem 1.2 it is sufficient to show that the family of Pfaff equations (ω_t) satisfies the assumptions of Theorem 1.1.

The equality (6.4) allows us to use Proposition 6.1 to conclude that the relation (6.2) holds for the path (6.5). It is clear that (6.2) implies that the family (ω_t) satisfies the assumptions (a) and (b) of Theorem 1.1 provided that the functions A_t in (6.2), $t \in [0, 1]$, vanish at no point of M^n . This will follow from the assumption on the C^∞ -closeness of ω_1 to ω_0 and Theorem B1 in Appendix B. Define a characteristic vector field X_t of (ω_t) by the relation $X_t \rfloor \Omega = \omega_t \wedge (d\omega_t)^{k-1}$, where Ω is a volume form. By (6.2) we have $X_t = A_t X_0$. The C^∞ -closeness of ω_1 to ω_0 implies the C^∞ -closeness of X_t , $t \in [0, 1]$, to X_0 . By Theorem B1 the C^∞ -closeness of X_t to X_0 in the equality $X_t = A_t X_0$ implies that the function A_t , $t \in [0, 1]$, is C^∞ -close to 1. Consequently, A_t vanishes at no point of the manifold. The proof of Theorem 1.2 is complete. \square

Proof of Proposition 6.1. Using (6.1), we may assume that the characteristic vector fields X_0 and X_1 of (ω_0) and (ω_1) , respectively, are equal. We shall prove that

$$(6.6) \quad X_0 \rfloor (\omega_t \wedge (d\omega_t)^{k-1}) = 0.$$

Having (6.6), we can use assumption (A) and the division property in Proposition 3.1.b (with polynomial dependence in t), which gives $\omega_t \wedge (d\omega_t)^{k-1} = X_0 \rfloor \mu_t$, where μ_t is a volume form. Let $\omega_0 \wedge (d\omega_0)^k = X_0 \rfloor \Omega$, $\mu_t = A_t \Omega$. Then we get (6.2).

To prove (6.6), we note that by Lemma 4.5 we have

$$(6.7) \quad X_0 \rfloor \omega_0 = X_0 \rfloor \omega_1 = 0;$$

therefore $X_0 \rfloor \omega_t = 0$. It follows that in order to prove (6.6) it suffices to prove the equality

$$(6.8) \quad (X_0 \rfloor d\omega_t) \wedge \omega_t = 0.$$

It is enough to prove the equality (6.8) at any point p such that $X_0(p) \neq 0$. At such a point $\omega_0(p) \neq 0$ and, since $X_0 = X_1$, $\omega_1(p) \neq 0$. From Lemma 4.5 we have

$(X_0 \rfloor d\omega_i) \wedge \omega_i = 0$, $i = 0, 1$. Thus, there are functions h_0 and h_1 , defined in a neighbourhood of p , such that in this neighbourhood we have

$$(6.9) \quad X_0 \rfloor d\omega_0 = h_0 \omega_0, \quad X_0 \rfloor d\omega_1 = h_1 \omega_1.$$

We will prove below that

$$(6.10) \quad h_1 = h_0.$$

Then from (6.9) we get $X_0 \rfloor d\omega_t = h\omega_t$, where $h = h_0 = h_1$, and so (6.8) holds.

We will show that (6.10) follows from (6.1). We take the Lie derivative of both parts in (6.1) along the vector field X_0 . Using the formula $L_X \eta = d(X \rfloor \eta) + X \rfloor d\eta$ for the Lie derivative, we obtain

$$\begin{aligned} L_{X_0} (\omega_0 \wedge (d\omega_0)^{k-1}) &= d(X_0 \rfloor (\omega_0 \wedge (d\omega_0)^{k-1})) + X_0 \rfloor (d\omega_0)^k \\ &= 0 + k(X_0 \rfloor d\omega_0) \wedge (d\omega_0)^{k-1} = kh_0 \omega_0 \wedge (d\omega_0)^{k-1}, \end{aligned}$$

and similarly

$$L_{X_0} (\omega_1 \wedge (d\omega_1)^{k-1}) = kh_1 \omega_1 \wedge (d\omega_1)^{k-1}.$$

Comparing these equalities and using (6.1) again, we get the required relation (6.10) (since $\omega_0 \wedge (d\omega_0)^{k-1} = \omega_1 \wedge (d\omega_1)^{k-1}$ does not vanish on a dense subset of M). Proposition 6.1 is proved. \square

7. PROOF OF THEOREM 1.3

Since the Martinet hypersurfaces of $P_t = (\omega_t)$ are the same for all t , the Martinet ideals are the same by the property of zeros. Thus we can fix a generator H of these ideals. The following two propositions will hold under the assumptions of Theorem 1.3. In the propositions all families are regular with respect to t (smooth in the C^∞ category and analytic in the C^ω category).

Proposition 7.1. *There exists a family of vector fields Y_t satisfying the relation*

$$(7.1) \quad \left(L_{Y_t} \omega_t + \frac{d\omega_t}{dt} \right) \wedge \omega_t \wedge (d\omega_t)^{k-1} = 0 \quad \text{mod } (H).$$

Proposition 7.2. *Let μ_t be a family of 1-forms such that*

$$(7.2) \quad \mu_t \wedge \omega_t \wedge (d\omega_t)^{k-1} = 0 \quad \text{mod } (H).$$

Then the equation

$$(7.3) \quad L_{Z_t} \omega_t + h_t \omega_t = \mu_t$$

has a solution (Z_t, h_t) .

The solvability of the homotopy equation (HE) in Section 1 is a direct corollary of these propositions. Namely, we take $\mu_t = -L_{Y_t} \omega_t - d\omega_t/dt$, and then the pair (\tilde{Z}_t, h_t) , with $\tilde{Z}_t = Z_t + Y_t$, solves the homotopy equation (HE).

Proof of Proposition 7.1. We fix a volume form Ω and define a family X_t of characteristic vector fields of (ω_t) by the relation $X_t \rfloor \Omega = \omega_t \wedge (d\omega_t)^{k-1} \wedge dH$. From Proposition 3.3 we have

$$X_t = R_t X_0 \quad \text{mod } (H),$$

where R_t is a family of nonvanishing functions, regular in t (of the same regularity in t as in the family ω_t). By Lemma 4.6 we have $X_t \rfloor \omega_t = 0 \quad \text{mod } (H)$ and $(X_t \rfloor d\omega_t) \wedge \omega_t = 0 \quad \text{mod } (H)$. We may replace X_t with X_0 in these equalities. In

particular, we get $X_0 \rfloor \omega_t = 0 \bmod (H)$, which implies $X_0 \rfloor (d\omega_t/dt) = 0 \bmod (H)$. Taking all these equalities into account, we see that

$$(7.4) \quad X_0 \rfloor \left(\frac{d\omega_t}{dt} \wedge \omega_t \wedge (d\omega_t)^{k-1} \right) = 0 \bmod (H).$$

This equality and Proposition 3.1.b imply that

$$\frac{d\omega_t}{dt} \wedge \omega_t \wedge (d\omega_t)^{k-1} = X_0 \rfloor (f_t \Omega) \bmod (H),$$

where f_t is a family of functions, regular in t . Replacing X_0 with $R_t^{-1}X_t$ and using the definition of X_t , we see that we can rewrite this relation in the form

$$\frac{d\omega_t}{dt} \wedge \omega_t \wedge (d\omega_t)^{k-1} = g_t \omega_t \wedge (d\omega_t)^{k-1} \wedge dH \bmod (H),$$

where $g_t = f_t/R_t$. This allows us to rewrite equation (7.1) in the form

$$(7.5) \quad (L_{Y_t} \omega_t - g_t dH) \wedge \omega_t \wedge (d\omega_t)^{k-1} = 0 \bmod (H).$$

It is clear that (7.5) holds if Y_t satisfies the relations

$$(7.6) \quad Y_t \rfloor d\omega_t = 0, \quad Y_t \rfloor \omega_t = g_t H,$$

since in this case

$$L_{Y_t} \omega_t = d(Y_t \rfloor \omega_t) = d(g_t H) = g_t dH \bmod (H).$$

Since (H) is the Martinet ideal of (ω_t) , we have

$$(7.7) \quad \omega_t \wedge (d\omega_t)^k = S_t H \Omega$$

for a family S_t of nonvanishing functions which has the same regularity in t as in ω_t (this follows from the regularity of the left-hand side and the fact that division by H is a continuous linear operator in the space of smooth functions, see Theorem B1 in Appendix B). Let us show that (7.6) holds for the family Y_t defined by

$$(7.8) \quad Y_t \rfloor \Omega = \frac{g_t}{S_t} (d\omega_t)^k.$$

In fact, applying $Y_t \rfloor$ to (7.8), we get $g_t (Y_t \rfloor d\omega_t) \wedge (d\omega_t)^{k-1} = 0$. This relation implies $Y_t \rfloor d\omega_t = 0$ (at points where $g_t(p) = 0$ we have $Y_t(p) = 0$, and at other points we can use Lemma 4.2 with $\lambda = (Y_t \rfloor d\omega_t)$ and the fact that contact points are dense). We have shown the first equality in (7.6). In order to prove the second one we apply $Y_t \rfloor$ to (7.7) and, using (7.8), we obtain that $(Y_t \rfloor \omega_t - g_t H) \cdot (d\omega_t)^k = 0$. This implies that $Y_t \rfloor \omega_t - g_t H = 0$ at points where the form $(d\omega_t)^k$ does not vanish, in particular, at points where ω_t is contact. By Proposition 3.2, (ii) the set of such points is everywhere dense; therefore $Y_t \rfloor \omega_t - g_t H = 0$ everywhere, and so (7.6) holds. Proposition 7.1 is proved. \square

Proof of Proposition 7.2. By Lemma 4.2, (i) and the fact that the set of contact points is dense in M^n (Proposition 3.2, (ii)), the equation (7.3) reduces to the following two equations:

$$(7.9) \quad (L_{Z_t} \omega_t) \wedge \omega_t \wedge (d\omega_t)^{k-1} = \mu_t \wedge \omega_t \wedge (d\omega_t)^{k-1},$$

$$(7.10) \quad (L_{Z_t} \omega_t + h_t \omega_t) \wedge (d\omega_t)^k = \mu_t \wedge (d\omega_t)^k$$

(with unknown Z_t and h_t), obtained from (7.3) by external multiplication by the forms $\omega_t \wedge (d\omega_t)^{k-1}$ and $(d\omega_t)^k$, respectively.

To solve equation (7.9) we use assumption (7.2). By this assumption

$$(7.11) \quad \mu_t \wedge \omega_t \wedge (d\omega_t)^{k-1} = H\nu_t$$

for some family ν_t of $2k$ -forms, regular in t by Theorem B1 in Appendix B. This permits us to find an explicit solution Z_t of (7.9). Namely, since H is a generator of the Martinet ideal of (ω_t) , we have relation (7.7), i.e., $\omega_t \wedge (d\omega_t)^k = HS_t\Omega$, where S_t is a family of nonvanishing functions, regular in t . Let us show that the family of vector fields Z_t defined by the relation

$$(7.12) \quad Z_t \rfloor \Omega = \frac{k}{S_t} \nu_t$$

is a solution of the equation (7.9). Relation (7.11) and the fact that H is not a zero divisor imply that $\nu_t \wedge \omega_t = 0$. This and (7.12) imply that

$$Z_t \rfloor \omega_t = 0.$$

Consequently, $L_{Z_t}\omega_t = Z_t \rfloor d\omega_t$ and

$$(7.13) \quad (L_{Z_t}\omega_t) \wedge \omega_t \wedge (d\omega_t)^{k-1} = (Z_t \rfloor d\omega_t) \wedge \omega_t \wedge (d\omega_t)^{k-1} = \frac{1}{k} Z_t \rfloor (\omega_t \wedge (d\omega_t)^k).$$

Now (7.9) follows from equalities (7.7) and (7.11)–(7.13).

To prove Proposition 7.2 it remains to solve equation (7.10) with respect to h_t . Since $Z_t \rfloor \omega_t = 0$, then $(L_{Z_t}\omega_t) \wedge (d\omega_t)^k = (k+1)^{-1} Z_t \rfloor (d\omega_t)^{k+1} = 0$, and the equation (7.10) takes the form

$$h_t \omega_t \wedge (d\omega_t)^k = \mu_t \wedge (d\omega_t)^k.$$

Due to relation (7.7), to prove that this equation has a solution h_t it suffices to prove that $\mu_t \wedge (d\omega_t)^k = HC_t\Omega$, where C_t is a family of functions, regular in t . We shall first prove that

$$(7.14) \quad (\mu_t \wedge (d\omega_t)^k)(p) = 0, \quad \text{for } p \in S.$$

This follows from relation (7.2). Namely, since ω_t is not contact at a point $p \in S$, thus (7.14) follows from (7.2) by Lemma 4.3, provided that $\omega_t(p) \neq 0$. Since the set of points of S at which ω_t vanishes is a subset of the set $\text{Sing}(L)$, the set of points $p \in S$ where $\omega_t(p) \neq 0$ is dense in S by Proposition 3.2, (iii). Therefore (7.14) holds at all points $p \in S$. By the property of zeros of the ideal (H) we obtain $\mu_t \wedge (d\omega_t)^k = HC_t\Omega$. Since μ_t and ω_t are regular in t , we deduce from Theorem B1 in Appendix B that C_t is regular in t . Proposition 7.2 is proved. \square

We have completed the proof of Theorem 1.3. Note that the extension property of S was used only when referring to Proposition 3.1.b., and therefore it is not needed if ω_t is polynomial in t (cf. the remark after Theorem 1.3).

8. PROOF OF THEOREM 1.4

Since the Martinet hypersurfaces for $P_0 = (\omega_0)$ and $P_1 = (\omega_1)$ are the same and the Martinet ideals have the property of zeros, they are equal and we can choose a common generator H which will be used throughout the proofs. The following proposition holds under the assumptions of Theorem 1.4 and will enable us to reduce the problem to Theorem 1.3.

Proposition 8.1. *Assume that*

$$\omega_1 \wedge (d\omega_1)^{k-1} \wedge dH = \omega_0 \wedge (d\omega_0)^{k-1} \wedge dH \quad \text{mod } (H).$$

Then, for the path $\omega_t = (1-t)\omega_0 + t\omega_1$ we have

$$(8.1) \quad \omega_t \wedge (d\omega_t)^{k-1} \wedge dH = B_t \omega_0 \wedge (d\omega_0)^{k-1} \wedge dH \quad \text{mod } (H),$$

and

$$(8.2) \quad \omega_t \wedge (d\omega_t)^k = C_t \omega_0 \wedge (d\omega_0)^k,$$

where B_t and C_t are families of functions, polynomial in t .

Proof of Theorem 1.4. Let X_0 and X_1 be characteristic vector fields of P_0 and P_1 defined via the same volume form and the same generator H of the Martinet ideal. Since $L(\omega_1) = L(\omega_0)$, then $X_1 \wedge X_0 = 0 \text{ mod } (H)$. From condition (A) and the division property in Proposition 3.1.a we obtain $X_1 = RX_0 \text{ mod } (H)$ or, equivalently,

$$(8.3) \quad \omega_1 \wedge (d\omega_1)^{k-1} \wedge dH = R\omega_0 \wedge (d\omega_0)^{k-1} \wedge dH \quad \text{mod } (H),$$

where R is a smooth or analytic function. We will later show, using closeness of ω_1 to ω_0 , that R is positive valued. Therefore, we can change the generator of P_1 for

$$(8.4) \quad \hat{\omega}_1 = R^{1/k} \omega_1,$$

and we get

$$(8.5) \quad \hat{\omega}_1 \wedge (d\hat{\omega}_1)^{k-1} \wedge dH = \omega_0 \wedge (d\omega_0)^{k-1} \wedge dH \quad \text{mod } (H).$$

Let

$$(8.6) \quad \omega_t = (1-t)\omega_0 + t\hat{\omega}_1.$$

To prove Theorem 1.4 it is enough to show that the family of Pfaff equations (ω_t) satisfies the assumptions of Theorem 1.3. Note that we do not need the extension property of S , since the family ω_t in (8.6) is polynomial (in fact, affine) in t , and in this case Theorem 1.3 was proved without using this assumption.

The equality (8.5) enables us to use Proposition 8.1. It is clear that relations (8.1) and (8.2) imply that the family (ω_t) satisfies the assumptions (a), (b) and (c) of Theorem 1.3 provided that the functions B_t do not vanish on S (then the characteristic line field does not change) and C_t do not vanish on M^n (then the Martinet ideal does not change), for $t \in [0, 1]$.

The fact that B_t and C_t do not vanish follows from the C^∞ -closeness of ω_1 to ω_0 and Theorem B1 in Appendix B. Since ω_1 is C^∞ -close to ω_0 , then the vector field X_1 is C^∞ close to X_0 . The relation $X_1 = RX_0 \text{ mod } (H)$ and continuity of the inverse to the operator $L_{X,H}$ in Theorem B1 imply that there exists a function \tilde{R} that is C^∞ -close to 1 and equal to R at any point of the Martinet hypersurface S . By the property of zeros of the Martinet ideal, $\tilde{R} = R \text{ mod } (H)$. We may replace R by \tilde{R} in (8.3) and in the definition (8.4) of $\hat{\omega}_1$. Then $\omega_t, t \in [0, 1]$, is C^∞ -close to ω_0 . Define a family of characteristic vector fields X_t by the relation $X_t|_\Omega = \omega_t \wedge (d\omega_t)^{k-1} \wedge dH$. Then $X_t, t \in [0, 1]$, is C^∞ -close to X_0 . The equality (8.1) is equivalent to $X_t = B_t X_0 \text{ mod } (H)$. We again use continuity of the inverse to the operator $L_{X,H}$ in Theorem B1. By this theorem there exists a function \tilde{B}_t that is C^∞ -close to 1 and equal to B_t at any point of S . Therefore $B_t > 0$ at any point of S .

To prove that C_t vanishes at no points of M^n , we also use the C^∞ -closeness of $\omega_t, t \in [0, 1]$, to ω_0 shown above. Let $\omega_t \wedge (d\omega_t)^k = Q_t H \Omega$, where Ω is a volume form. Relation (8.2) implies that $Q_t = C_t Q_0$. The C^∞ -closeness of ω_t to ω_0 implies the C^∞ -closeness of the function HQ_t to HQ_0 . By continuity of the inverse to the operator $f \rightarrow fH$ (Theorem B1) we get the C^∞ -closeness of Q_t to Q_0 . Since Q_0 is a nonvanishing function, then C_t is C^∞ -close to 1. The proof of Theorem 1.4 is complete. \square

Proof of Proposition 8.1. It is enough to prove the equalities

$$(8.7) \quad X_0 \rfloor (\omega_t \wedge (d\omega_t)^{k-1} \wedge dH) = 0 \pmod{(H)},$$

$$(8.8) \quad X_0 \rfloor (\omega_t \wedge (d\omega_t)^k) = 0 \pmod{(H)}.$$

Namely, equality (8.7) and condition (A) allow us to use the division properties in Proposition 3.1.b to conclude that $\omega_t \wedge (d\omega_t)^{k-1} \wedge dH = B_t X_0 \rfloor \Omega \pmod{(H)}$, and so the validity of relation (8.1), where B_t is a family of functions, polynomial in t . Equality (8.8) implies that $\omega_t \wedge (d\omega_t)^k$ vanishes at those points of S at which X_0 does not vanish. By Proposition 3.2, (iii) the set of such points is dense in S ; therefore $\omega_t \wedge (d\omega_t)^k$ vanishes at all points of S . By the property of zeros of the Martinet ideal we have $\omega_t \wedge (d\omega_t)^k = 0 \pmod{(H)}$, and consequently (8.2) holds for some family of functions C_t , polynomial in t .

In order to prove (8.7) and (8.8) we use the assumption of Proposition 8.1 and choose characteristic vector fields X_0 and X_1 of (ω_0) and (ω_1) equal modulo (H) . By Lemma 4.6 we have

$$X_0 \rfloor \omega_i = 0 \pmod{(H)}, \quad X_0 \rfloor dH = 0 \pmod{(H)},$$

for $i = 0, 1$, and therefore $X_0 \rfloor \omega_t = 0 \pmod{(H)}$. It follows that in order to prove (8.7) and (8.8) it suffices to prove the equality

$$(8.9) \quad (X_0 \rfloor d\omega_t) \wedge \omega_t = 0 \pmod{(H)}.$$

Due to the property of zeros of the Martinet ideal, it suffices to prove this equality for any point $p \in S$ such that $X_0(p) \neq 0$. At such points the 1-form dH does not vanish and S is smooth. Since $X_1 = X_0 \pmod{(H)}$, using Lemma 4.6 we obtain

$$(X_0 \rfloor d\omega_0) \wedge \omega_0 = (X_0 \rfloor d\omega_1) \wedge \omega_1 = 0$$

in a neighbourhood of p in S . Since $X_0(p) \neq 0$, then $\omega_0(p) \neq 0$ and $\omega_1(p) \neq 0$, and therefore these relations imply the equalities

$$(8.10) \quad X_0 \rfloor d\omega_0 = h_0 \omega_0, \quad X_0 \rfloor d\omega_1 = h_1 \omega_1,$$

which hold in a neighbourhood U of p in S . Here h_0 and h_1 are functions defined in this neighbourhood. We will show that $h_0 = h_1$; then (8.10) implies that $X_0 \rfloor d\omega_t = h_0 \omega_t$, and (8.9) holds in the neighbourhood U .

To prove that $h_0 = h_1$ on $U \subset S$, we restrict the relation assumed in Proposition 8.1 to the tangent bundle of U . We obtain $(\omega_1 \wedge d\omega_1)|_U = (\omega_0 \wedge (d\omega_0)^{k-1})|_U$ and take the Lie derivative of this relation along the restriction $X_0|_S$ of X_0 to S (recall that X_0 is tangent to S). As in the proof of Proposition 6.1, we obtain the required equality $h_0 = h_1$. Proposition 8.1 is proved. \square

APPENDIX A. DIVISION PROPERTIES

In this Appendix we present a general theorem on division properties of the exterior (respectively, interior) product with a section X of a vector bundle. This theorem is proved in [DJ] and implies our Propositions 3.1.a and 3.1.b. Our results hold in the categories C^s , where $s = \infty$ or $s = \omega$.

Let M be a paracompact differential manifold. Consider a vector bundle E over M of rank m and denote by E^* its dual bundle. Let $\Lambda_r = \Lambda_r(E)$ denote the r th exterior power of E , $r = 0, 1, \dots, m$, with $\Lambda_0 = M \times \mathbb{R}$ and $\Lambda_1 = E$. We denote by $\Lambda_r(M; E)$ the linear space of sections of Λ_r (smooth or real analytic, depending on the category).

Any section α of E defines the linear operator of exterior multiplication by α , which gives the complex

(A.1)
$$0 \rightarrow \Lambda_0(M) \rightarrow \Lambda_1(M; E) \rightarrow \dots \rightarrow \Lambda_m(M; E),$$

with the operator $\partial_\alpha = \partial_\alpha^p : \Lambda_p(M; E) \rightarrow \Lambda_{p+1}(M; E)$ defined by $\partial_\alpha^p(\gamma) := \alpha \wedge \gamma$.

Consider a section X of the dual bundle E^* . This section defines the operator of the interior product with X , $X] : \Lambda_r(M; E) \rightarrow \Lambda_{r-1}(M; E)$. Given a local basis e_1, \dots, e_m of E , the operator of the interior product with X is defined on the elements of a local basis of Λ_r by

$$X](e_{i_1} \wedge \dots \wedge e_{i_m}) = \sum_{j=1}^m (-1)^{j-1} \langle X, e_{i_j} \rangle e_{i_1} \wedge \dots \check{e}_{i_j} \dots \wedge e_{i_m},$$

where \check{e}_{i_j} means absence of e_{i_j} and $\langle \cdot, \cdot \rangle$ denotes the duality product between E^* and E . Clearly, $(X])^2 = 0$; so the operator $X]$ defines the complex

(A.2)
$$0 \rightarrow \Lambda_m(M; E) \rightarrow \Lambda_{m-1}(M; E) \rightarrow \dots \rightarrow \Lambda_1(M; E) \rightarrow \Lambda_0(M; E).$$

Let S be a closed subset of M . Denote by $\Lambda_r(M, S; E) \subset \Lambda_r(M; E)$ the subspace of sections of $\Lambda_r(M; E)$ vanishing at all points of S , and let

$$\Lambda_r(S; E) = \Lambda_r(M; E) / \Lambda_r(M, S; E)$$

denote the quotient space.

Any element α of $\Lambda_1(S; E)$ defines the unique operator

$$\partial_\alpha^p : \Lambda_p(S; E) \rightarrow \Lambda_{p+1}(S; E)$$

(the quotient of the operator of exterior multiplication), which gives the complex

(A.3)
$$0 \rightarrow \Lambda_0(S) \rightarrow \Lambda_1(S; E) \rightarrow \dots \rightarrow \Lambda_m(S; E).$$

Given a section X of E^* , the operator $X]$ defines the following complex on the quotient spaces:

(A.4)
$$0 \rightarrow \Lambda_m(S; E) \rightarrow \Lambda_{m-1}(S; E) \rightarrow \dots \rightarrow \Lambda_1(S; E) \rightarrow \Lambda_0(S; E).$$

We define the invariant $d_p(X) = \text{depth}(I_p)$, where I_p is the ideal of function germs at $p \in M$ generated by the coefficients a_1, \dots, a_m of X in a local basis of E^* (cf. Definition 1.2 in Section 1). Similarly, given a pair (H, X) of a function H and a section X of E^* on M , we define $d_p(H, X)$ as the maximal length of a regular sequence of function germs that begins with the germ H_p of H at p and has further elements in I_p (cf. Definition 1.4). Analogously we define the invariants $d_p(\alpha)$ and $d_p(H, \alpha)$.

Statements (i) and (ii) of the following theorem hold in the C^∞ and C^ω categories, for $0 \leq q \leq n-1$.

Theorem A. (i) If α satisfies the condition $d_p(\alpha) \geq q+1$ for all $p \in M$ such that $\alpha(p) = 0$, then the complex (A.1) is exact up to $\Lambda_q(M; E)$. Similarly, if $d_p(X) \geq q+1$ for all $p \in M$ such that $X(p) = 0$, then the complex (A.2) is exact up to $\Lambda_{m-q}(M; E)$.

(ii) Let H be a function on M such that the ideal (H) has the property of zeros, and let $S = \{H = 0\}$. If α is a section of E on M such that (H, α) satisfies $d_p(H, \alpha) \geq q+2$ for all $p \in S$ such that $\alpha(p) = 0$, then the complex (A.3) is exact up to $\Lambda_q(S; E)$. Similarly, if X is a section of E^* on M and $d_p(H, X) \geq q+2$ for all $p \in S$ such that $X(p) = 0$, then the complex (A.4) is exact up to $\Lambda_{m-q}(S; E)$.

(iii) If the assumptions of (i) hold, then, in the C^∞ category, the complex (A.1) splits up to $\Lambda_{q-1}(M; E)$ and the complex (A.2) splits up to $\Lambda_{m-q+1}(M; E)$. If the assumptions of (ii) hold and S has the extension property, then the complex (A.3) splits up to $\Lambda_{q-1}(S; E)$ and the complex (A.4) splits up to $\Lambda_{m-q+1}(S; E)$. Here the corresponding spaces are equipped with the C^∞ topology and are considered as Fréchet spaces (quotient Fréchet spaces).

Above, a complex $0 \rightarrow L_m \rightarrow \cdots \rightarrow L_{m-q+1} \rightarrow L_{m-q} \rightarrow \cdots$ defined by the operators $\partial_i : L_i \rightarrow L_{i-1}$ is called exact up to L_{m-q} if $\text{Im } \partial_{i+1} = \ker \partial_i$ for $i = m, m-1, \dots, m-q$, and it splits up to L_{m-q+1} if the L_i are linear topological spaces, $\text{Im } \partial_i$ are closed subspaces of L_{i-1} and each $\partial_i : L_i \rightarrow L_{i-1}$ has a continuous right inverse K_i defined on $\text{Im } \partial_i$, for all $i = m, m-1, \dots, m-q+1$.

The above theorem follows from Theorems 2.1 and 2.2 in [DJ]. In the local case (of germs) statements (i) and (ii) follow from a well-known algebraic result on exactness of the Koszul complex, cf. e.g. [E] or [JZh2], Appendix 1.

Proof of Proposition 3.1.a. In the even-dimensional case the first implication follows trivially from statement (i) in Theorem A if we take the bundle E equal to the cotangent bundle $E = T^*M$, the dual $E^* = TM$, and consider the complex (A.2). ($\Lambda_r(M, E)$ is identified with the space of differential r -forms on M .) The second implication follows analogously from the same statement by taking $E = TM$ and the complex (A.1).

In the odd-dimensional case the first implication follows in a similar way from statement (ii) in Theorem A concerning the complex (A.4). This is because the property of zeros of (H) allows us to identify the elements of $\Lambda_r(S; E)$ with the equivalence classes of differential r -forms modulo (H) (cf. our convention on notation mod (H) presented after Definition 1.3). The second implication follows analogously from statement (ii) in Theorem A concerning the complex (A.3). \square

Proof of Proposition 3.1.b. The existence of μ_t and f_t for any fixed t follows from Proposition 3.1.a. We have to show the regularity of these families in t . We shall prove the regularity of μ_t (the proof of regularity of f_t is analogous). If ν_t depends on t polynomially, then Proposition 3.1 allows us to construct μ_t polynomial in t , and the regularity follows trivially. In the general case our arguments are different for the categories C^ω and C^∞ .

In the C^ω category we use the following fact: if a sequence a_1, \dots, a_r of real analytic function germs at $p \in M$ is regular in the ring of analytic function germs at p , then it is regular when considered as a sequence in the ring of real analytic function germs of the variables $(x, t) \in M \times \mathbb{R}$ at (p, t_0) , for any $t_0 \in [0, 1]$. Using

local coordinates, this fact can be easily proved for the case of formal power series using the definition of regular sequence in the ring of formal power series of the variables x_1, \dots, x_n, t . Then, using the fact that the ring of formal power series is faithfully flat over the ring of convergent series (see Malgrange [Mlg], Chapter 3), we see that it also holds for converging series and so for germs of analytic functions. Using the above fact we see that the assumption (A) holds over the manifold $\tilde{M} = M \times I$, where I is an open interval containing $[0, 1]$ on which the analytic family ν_t is well defined by analytic extension. Thus we can use Theorem A over the manifold \tilde{M} , i.e., for the bundles $E = T^*M$ and $E^* = TM$ pulled back to \tilde{M} by the canonical projection $M \times I \rightarrow M$.

In the C^∞ category, in the even-dimensional case the smooth dependence of μ_t on t follows from statement (iii) in Theorem A. By this statement there exists a continuous right inverse operator K to the linear operator $X] : \Lambda_{r+1}(M; E) \rightarrow \Lambda_r(M; E)$, for $r = n-1$ and $r = n-2$, and we can define $\mu_t = K\nu_t$. Here $E = T^*M$ and $\Lambda_r(M; E) = \Lambda_r(M)$, the space of differential r -forms on M .

In the C^∞ category, in the odd-dimensional case we also use statement (iii) of Theorem A and the extension property of S . Namely, for $E = T^*M$ we define $\mu_t = \lambda K\nu_t|_S$, where $K : \Lambda_{n-1}(S; E) \rightarrow \Lambda_n(S; E)$ is the continuous right inverse operator to $X] : \Lambda_n(S; E) \rightarrow \Lambda_{n-1}(S; E)$, and $\lambda : \Lambda_n(S; E) \rightarrow \Lambda_n(M; E)$ is a continuous linear operator of extension. \square

APPENDIX B. CONTINUITY OF DIVISION

Continuity of division in the cases presented below is needed in the main proofs and will be proved separately. Let $C^\infty(M)$ and $Vect^\infty(M)$ be the spaces of smooth functions and smooth vector fields on M , with the C^∞ topology. Let $C^\infty(M, S)$ and $Vect^\infty(M, S)$ be the subspaces of functions (vector fields) on M vanishing on the Martinet hypersurface $S \subset M$. The quotient Fréchet spaces

$$C^\infty(S) = C^\infty(M)/C^\infty(M, S), \quad Vect^\infty(S; TM) = Vect^\infty(M)/Vect^\infty(M, S)$$

can be identified with the space of smooth functions on S and the space of smooth sections of the tangent bundle TM restricted to S , respectively.

Given a Pfaff equation on M^{2k} and a characteristic vector field X , we consider the linear operator

$$L_X : C^\infty(M) \rightarrow Vect^\infty(M), \quad L_X(f) = fX.$$

For a Pfaff equation on M^{2k+1} , a characteristic vector field X and a generator H of the Martinet ideal we consider the linear operators

$$L_H : C^\infty(M) \rightarrow C^\infty(M), \quad L_H(f) = Hf,$$

and

$$L_{X,H} : C^\infty(S) \rightarrow Vect^\infty(S; TM), \quad L_{X,H}([f]) = [fX],$$

where $[\cdot]$ denotes the equivalence class in the corresponding quotient space.

Theorem B1. *If the characteristic vector field X satisfies condition (A) and in the odd-dimensional case the Martinet ideal (H) has the property of zeros, then each of the linear operators L_X, L_H and $L_{X,H}$ is injective, has closed image, and has continuous inverse defined on the image.*

Proof. By the Banach open mapping theorem in Fréchet spaces it suffices to prove that each of the operators L_X , L_H , and $L_{X,H}$ is injective and has closed image.

The injectivity of the operators L_X , L_H and $L_{X,H}$ follows from Proposition 3.2, (i), (ii), (iii), respectively. The closedness of the image of the operator L_H follows from the global property of zeros implied by Definition 1.5—by this property the image of L_H coincides with the closed subspace $C^\infty(M, S) \subset C^\infty(M)$ of functions vanishing on S .

To prove the closedness of the image of the operators L_X and $L_{X,H}$, we use Proposition 3.1. By Proposition 3.1, (i) the image of L_X coincides with the kernel of the continuous operator $Vect^\infty(M) \rightarrow \Gamma^\infty(\Lambda^2 TM)$ defined by $Y \rightarrow X \wedge Y$, where $\Gamma^\infty(\Lambda^2 TM)$ is the space of smooth sections of the skew-symmetric product of the tangent bundle TM , with the C^∞ topology. Similarly, by Proposition 3.1, (ii) the image of $L_{X,H}$ coincides with the kernel of the continuous operator $Vect^\infty(S) \rightarrow \Gamma^\infty(\Lambda^2(S; TM))$ given by $[Y] \rightarrow [X \wedge Y]$, where $\Gamma^\infty(\Lambda^2(S; TM))$ is the space of smooth sections over S of the skew-symmetric product of the tangent bundle TM (with the C^∞ topology) and $[\cdot]$ denotes the equivalence class in the corresponding quotient space. The kernel of this operator is a closed subspace of $Vect^\infty(S; TM) = Vect^\infty(M)/Vect^\infty(M, S)$. The proof is complete. \square

It is natural to ask if it is possible to replace the C^∞ -closeness of ω_1 to ω_0 in Theorems 1.2 and 1.4 by C^r -closeness with some r . Any attempt at answering this question requires modification of Theorem B1, which was used in the proofs of Theorems 1.2 and 1.4. Proving Theorem 1.2, we had to show that the function A_t , $t \in [0, 1]$, does not vanish at any point of M . In the proof of Theorem 1.4 we had to show that the functions B_t , $t \in [0, 1]$, do not vanish at points of S and the functions C_t , $t \in [0, 1]$, do not vanish at points of M . The C^∞ -closeness of ω_1 to ω_0 given as an assumption in Theorems 1.2 and 1.4 and the continuity of the inverse to the operators L_X , L_H and $L_{X,H}$ allowed us to obtain the C^∞ -closeness of A_t , B_t and C_t to 1. Of course, to conclude that these functions do not vanish, their C^0 -closeness to 1 would be enough.

We introduce the following topological characteristic of a linear injective operator $L : C^\infty(M) \rightarrow C^\infty(M)$ or $L : C^\infty(M) \rightarrow Vect^\infty(M)$ or $L : C^\infty(S) \rightarrow Vect^\infty(S)$. Denote by $m \in \{0, 1, 2, \dots; \infty\}$ the minimal m such that for any $s \geq 0$ the convergence to 0 of the sequence of sections $L(f_n)$ in the C^{s+m} topology implies the convergence to 0 of the sequence of functions f_n in the C^s topology. This means that the inverse to L behaves not worse than a linear differential operator of order m . Note that by Theorem B1 we have $m(L_X)$, $m(L_H)$, and $m(L_{X,H}) \leq \infty$.

In many cases the numbers $m(L_X)$, $m(L_H)$ and $m(L_{X,H})$ are finite and can be found or estimated from above, see examples below. Tracing the construction of the functions A_t , B_t and C_t in the proofs of Theorems 1.2 and 1.4, it is easy to check that if these numbers are finite, then:

1. the C^0 -closeness of A_t to 1 holds provided that the 1-form ω_1 is close to ω_0 in the C^r topology with $r = 2m(L_X) + 2$;
2. the C^0 -closeness of B_t to 1 holds provided that the 1-form ω_1 is close to ω_0 in the C^r topology with $r = 2m(L_{X,H}) + 2$;
3. the C^0 -closeness of C_t to 1 holds provided that the 1-form ω_1 is close to ω_0 in the C^r topology with $r = m(L_{X,H}) + m(L_H) + 2$.

Therefore in Theorems 1.2 and 1.4 the C^∞ -closeness of ω_1 to ω_0 can be replaced by the closeness in a weaker topology, and we obtain the following result.

Theorem B2. In Theorem 1.2 the C^∞ -closeness of ω_1 to ω_0 can be replaced by the C^r -closeness with $r = 2m(L_X) + 2$. In Theorem 1.4 the C^∞ -closeness of ω_1 to ω_0 can be replaced by the C^r -closeness with

$$r = \max(2m(L_{X,H}) + 2, m(L_{X,H}) + m(L_H) + 2).$$

Examples ($n = 2k$). 1. If X has no singular points, then it is clear that $m(L_X) = 0$. Therefore the C^∞ -closeness of ω_1 to ω_0 in Theorem 1.2 can be replaced by C^2 -closeness. We obtain Theorem 0.1.

2. If the 1-jet of X vanishes at no points of the manifold, then it is easy to prove that $m(L_X) \leq 1$. Therefore the C^∞ -closeness of ω_1 to ω_0 in Theorem 1.2 can be replaced by C^4 -closeness.

Examples ($n = 2k + 1$). 1. If (ω_0) is a Martinet distribution, i.e., $dH(p) \neq 0$ and $X(p) \neq 0$ for any $p \in S = \{H = 0\}$, then it is easy to prove that $m(L_H) \leq 1$ and $m(L_{X,H}) = 0$. Therefore the C^∞ -closeness of ω_1 to ω_0 in Theorem 1.4 can be replaced by C^3 -closeness. We obtain Theorem 0.2.

2. Assume that $dH(p) \neq 0$ for any $p \in S = \{H = 0\}$. Then the restriction of X to S is a smooth vector field $X|_S$ on S . Assume that the 1-jet of X_S does not vanish. In this case $m(L_H) \leq 1$ and $m(L_{X,H}) \leq 1$. Therefore the C^∞ -closeness of ω_1 to ω_0 in Theorem 1.4 can be replaced by C^4 -closeness.

ACKNOWLEDGMENTS

While working on this paper we have profited from discussions with several colleagues. We are especially thankful for helpful advice obtained from Paweł Domański, Jean-Paul Gauthier, Pierre Milman and Richard Montgomery.

REFERENCES

- [A] A. Agrachev, Methods of Control Theory in Nonholonomic Geometry, Proc. Internat. Congress of Math., Zürich 1994, Vol. 2, pp. 1473-1483, Birkhäuser, Basel, 1995. MR **97f**:58051
- [ArII] V. I. Arnold and Yu. S. Ilyashenko, *Ordinary Differential Equations*, in Encyclopaedia of Math. Sci. Vol. 1, Dynamical Systems 1, Springer-Verlag (1986). MR **87e**:34049; MR **89g**:58060
- [BS] E. Bierstone and G. W. Schwarz, Continuous linear division and extension of C^∞ functions, *Duke Math. Journal* 50 (1983), 233-271. MR **86b**:32010
- [BH] R. L. Bryant and L. Hsu, Rigidity of integral curves of rank 2 distributions, *Inventiones Math.* 114 (1993), 435-461. MR **94j**:58003
- [C] H. Cartan, Variétés analytiques réelles et variétés analytiques complexes, *Bull. Soc. Math. de France* 85 (1957), 77-99. MR **20**:1339
- [DJ] P. Domański and B. Jakubczyk, Linear continuous division for exterior and interior products (accepted to Proc. Amer. Math. Soc.).
- [E] D. Eisenbud, *Commutative Algebra*, Springer-Verlag, New York 1994. MR **97a**:13001
- [Gol] A. Golubev, On the global stability of maximally nonholonomic two-plane fields in four dimensions, *Internat. Math. Res. Notices*, 11, 523 - 529, 1997. MR **98g**:57043
- [G] J. W. Gray, Some global properties of contact structures, *Annals of Math.* 69, (1959), 421-450. MR **22**:3016
- [JP] B. Jakubczyk and F. Przytycki, On J. Martinet's conjecture, *Bull. Polish Acad. Sci. Math.* 27 (1979), No. 9, 731-735. MR **82e**:58003
- [JZh1] B. Jakubczyk and M. Zhitomirskii, Odd-dimensional Pfaffian equations; reduction to the hypersurface of singular points, *Comptes Rendus Acad. Sci. Paris, Série I* t. 325 (1997), 423-428. MR **99e**:58003
- [JZh2] B. Jakubczyk and M. Zhitomirskii, Local reduction theorems and invariants for singular contact structures, *Ann. Inst. Fourier*, 51 (2001), 237-295. MR **2002c**:58001

- [LS] W. Liu and H. Sussmann, Shortest paths for sub-Riemannian metrics on rank-two distributions, *Mem. Amer. Math. Soc.*, Vol. 118 (1995), No. 564. MR **96c**:53061
- [Mlg] B. Malgrange, Ideals of differentiable functions, Oxford University Press, 1966. MR **35**:3446
- [Mar] J. Martinet, Sur les singularités des formes différentielles, *Ann. Inst. Fourier*, Vol. 20, No.1 (1970), 95-178. MR **44**:3333
- [Mon] R. Montgomery, A survey on singular curves in sub-Riemannian geometry, *J. Dynamical and Control Systems*, Vol.1, No.1 (1995), 49-90. MR **95m**:53060
- [MZh] R. Montgomery and M. Zhitomirskii, Geometric approach to Goursat flags, *Annales de l'Institut Henri Poincaré, Analyse Nonlineaire*, Vol. 18, No. 4 (2001), 459-493. MR **2002d**:58004
- [Ru] J. M. Ruiz, The Basic Theory of Power Series, *Advanced Lectures in Mathematics*, Vieweg, Wiesbaden, 1993. MR **94i**:13012
- [TEG] C. B. Thomas, Y. Eliashberg, and E. Giroux, 3-dimensional contact geometry, in "Contact and Symplectic Geometry", Publ. Newton Institute 8, Cambridge University Press, Cambridge, 1996, 48-65. MR **98b**:53026
- [Zh1] M. Zhitomirskii, Singularities and normal forms of odd-dimensional Pfaff equations. *Functional. Anal. Appl.*, Vol. 23, (1989), No. 1, pp. 59-61. MR **90i**:58007
- [Zh2] M. Zhitomirskii, Typical singularities of differential 1-forms and Pfaffian equations, *Translations of Math. Monographs*, Vol. 113, Amer. Math. Soc., Providence, RI, 1992. MR **94j**:58004

INSTITUTE OF MATHEMATICS, POLISH ACADEMY OF SCIENCES, ŚNIADECKICH 8, 00-950 WARSAW, POLAND AND INSTITUTE OF APPLIED MATHEMATICS, UNIVERSITY OF WARSAW, POLAND
E-mail address: B.Jakubczyk@impan.gov.pl

DEPARTMENT OF MATHEMATICS, TECHNION, 32000 HAIFA, ISRAEL
E-mail address: mzhi@techunix.technion.ac.il

WHEN ARE THE TANGENT SPHERE BUNDLES OF A RIEMANNIAN MANIFOLD REDUCIBLE?

E. BOECKX

ABSTRACT. We determine all Riemannian manifolds for which the tangent sphere bundles, equipped with the Sasaki metric, are local or global Riemannian product manifolds.

1. INTRODUCTION

When studying the geometry of a Riemannian manifold (M, g) , it is often useful to relate it to the properties of its unit tangent sphere bundle T_1M . In earlier work, we have been primarily interested in the geometric properties of T_1M when equipped with the Sasaki metric g_S . This is probably the simplest possible Riemannian metric on T_1M and it is completely determined by the metric g on the base manifold M . In this way, we have obtained a number of interesting characterizations of specific classes of Riemannian manifolds. We refer to [2], [5], [6], [7] and the references therein for examples of this. Also tangent sphere bundles T_rM with radius r different from 1 and equipped with the Sasaki metric have been studied recently ([9], [10]). The geometric properties of these Riemannian manifolds may change with the radius. See [9] for an example of this. Of course, other Riemannian metrics on the tangent bundle and on the tangent sphere bundles are possible. Of these, the Cheeger-Gromoll metric g_{CG} may be the best known. However, for tangent sphere bundles, this specific metric yields nothing new, since (T_rM, g_{CG}) is isometric to $(T_{r/\sqrt{1+r^2}}M, g_S)$. The isometry is given explicitly by $\phi: T_rM \rightarrow T_{r/\sqrt{1+r^2}}M: (x, u) \mapsto (x, u/\sqrt{1+r^2})$.

It is an interesting geometric problem to determine when a tangent sphere bundle, which we always consider with the Sasaki metric in this paper, is reducible, i.e., when it is locally or globally isometric to a Riemannian product manifold. To our surprise, we could not find any results in the literature concerning this question. Nevertheless, knowledge about reducibility could help to deal with geometric questions about tangent sphere bundles. In [4] for instance, we use it in an essential way to determine all unit tangent sphere bundles that are semi-symmetric, i.e., for which the curvature tensor at each point is algebraically the same as that of some symmetric space. Actually, that problem was the inspiration for the present article. As concerns the local reducibility of tangent sphere bundles, we prove here the following.

Received by the editors November 11, 2002 and, in revised form, January 21, 2003.

2000 *Mathematics Subject Classification*. 53B20, 53C12, 53C20.

Key words and phrases. Tangent sphere bundle, Sasaki metric, reducibility, Clifford structures, foliations.

Local Theorem. *A tangent sphere bundle $(T_r M, g_S)$, $r > 0$, of a Riemannian manifold (M^n, g) , $n \geq 2$, is locally reducible if and only if (M, g) has a flat factor, i.e., (M, g) is locally isometric to a product $(M', g') \times (\mathbb{R}^k, g_0)$ where $1 \leq k \leq n$ and g_0 denotes the standard Euclidean metric on \mathbb{R}^k .*

The corresponding global version reads as follows:

Global Theorem. *Let (M^n, g) , $n \geq 3$, be a Riemannian manifold and suppose that $(T_r M, g_S)$ is a global Riemannian product. Then, (M, g) is either flat or it is also a global Riemannian product, with a flat factor.*

Conversely, if (M, g) is a global product space $(M', g') \times (F^k, g_0)$ where $1 \leq k \leq n$ and F is a connected and simply connected flat space, then $(T_r M, g_S)$ is a global Riemannian product, also with (F, g_0) as a flat factor.

In view of the comments above, these results remain valid if we consider the tangent sphere bundles equipped with the Cheeger-Gromoll metric.

This article is organized as follows. After giving the necessary definitions and formulas concerning tangent sphere bundles, we show in Section 3 that only two types of decomposition for $T_r M$ are possible: a vertical and a diagonal one. The special form of the curvature of $(T_r M, g_S)$ for vertical vectors is crucial here. In particular, the same procedure does not go through for the tangent bundle TM . Section 4 deals with the diagonal case. We find that a diagonal decomposition gives rise to a Clifford representation via specific curvature operators. As a result, only base manifolds with dimension 2, 3, 4, 7 or 8 could possibly admit diagonal decompositions. The different dimensions are then handled separately. It turns out that diagonal decompositions can only be realized for a flat surface as base space. The general situation with a vertical decomposition is treated in Section 5 and leads to the Local Theorem above. The final section is devoted to global considerations.

2. TANGENT SPHERE BUNDLES

We first recall a few of the basic facts and formulas about the tangent sphere bundles of a Riemannian manifold. A more elaborate exposition and further references can be found in [5] and [9].

The tangent bundle TM of a Riemannian manifold (M, g) consists of pairs (x, u) where x is a point in M and u is a tangent vector to M at x . The mapping $\pi: TM \rightarrow M: (x, u) \mapsto x$ is the natural projection from TM onto M . It is well known that the tangent space to TM at a point (x, u) splits into the direct sum of the vertical subspace $VTM_{(x, u)} = \ker \pi_{*|(x, u)}$ and the horizontal subspace $HTM_{(x, u)}$ with respect to the Levi-Civita connection ∇ of (M, g) : $T_{(x, u)} TM = VTM_{(x, u)} \oplus HTM_{(x, u)}$.

For $w \in T_x M$, there exists a unique horizontal vector $w^h \in HTM_{(x, u)}$ for which $\pi_*(w^h) = w$. It is called the *horizontal lift* of w to (x, u) . There is also a unique vertical vector $w^v \in VTM_{(x, u)}$ for which $w^v(df) = w(f)$ for all functions f on M . It is called the *vertical lift* of w to (x, u) . These lifts define isomorphisms between $T_x M$ and $HTM_{(x, u)}$ and $VTM_{(x, u)}$, respectively. Hence, every tangent vector to TM at (x, u) can be written as the sum of a horizontal and a vertical lift of uniquely defined tangent vectors to M at x . The *horizontal* (respectively *vertical*) *lift* of a vector field X on M to TM is defined in the same way by lifting X pointwise. Further, if T is a tensor field of type $(1, s)$ on M and X_1, \dots, X_{s-1} are vector fields on M , then we denote by $T(X_1, \dots, u, \dots, X_{s-1})^v$ the vertical vector field on TM

which at (x, w) takes the value $T(X_{1x}, \dots, w, \dots, X_{s-1x})^v$, and similarly for the horizontal lift. In general, these are *not* the vertical or horizontal lifts of a vector field on M .

The *Sasaki metric* g_S on TM is completely determined by

$$g_S(X^h, Y^h) = g_S(X^v, Y^v) = g(X, Y) \circ \pi, \quad g_S(X^h, Y^v) = 0$$

for vector fields X and Y on M .

Our interest lies in the tangent sphere bundle $T_r M$ of some positive radius r , which is a hypersurface of TM consisting of all tangent vectors to (M, g) of length r . It is given implicitly by the equation $g_x(u, u) = r^2$. A unit normal vector field N to $T_r M$ is given by the vertical vector field u^v/r . We see that horizontal lifts to $(x, u) \in T_r M$ are tangent to $T_r M$, but vertical lifts in general are not. For that reason, we define the *tangential lift* w^t of $w \in T_x M$ to $(x, u) \in T_r M$ by $w^t = w^v - \frac{1}{r} g(w, u)N$. Clearly, the tangent space to $T_r M$ at (x, u) is spanned by horizontal and tangential lifts of tangent vectors to M at x . One defines the *tangential lift of a vector field* X on M in the obvious way. For the sake of notational clarity, we will use \bar{X} as a shorthand for $X - \frac{1}{r^2} g(X, u)u$. Then $X^t = \bar{X}^v$. Further, we denote by $VT_r M$ the $(n-1)$ -dimensional distribution of vertical tangent vectors to $T_r M$.

If we consider $T_r M$ with the metric induced from the Sasaki metric g_S of TM , also denoted by g_S , we turn $T_r M$ into a Riemannian manifold. Its Levi-Civita connection $\bar{\nabla}$ is described completely by

$$\begin{aligned} \bar{\nabla}_{X^t} Y^t &= -\frac{1}{r^2} g(Y, u) X^t, \\ \bar{\nabla}_{X^t} Y^h &= \frac{1}{2} (R(u, X) Y)^h, \\ \bar{\nabla}_{X^h} Y^t &= (\nabla_X Y)^t + \frac{1}{2} (R(u, Y) X)^h, \\ \bar{\nabla}_{X^h} Y^h &= (\nabla_X Y)^h - \frac{1}{2} (R(X, Y) u)^t \end{aligned} \quad (1)$$

for vector fields X and Y on M . Its Riemann curvature tensor \bar{R} is given by

$$\begin{aligned} \bar{R}(X^t, Y^t) Z^t &= \frac{1}{r^2} (g(\bar{Y}, \bar{Z}) X^t - g(\bar{Z}, \bar{X}) Y^t), \\ \bar{R}(X^t, Y^t) Z^h &= (R(\bar{X}, \bar{Y}) Z)^h + \frac{1}{4} ([R(u, X), R(u, Y)] Z)^h, \\ \bar{R}(X^h, Y^t) Z^t &= -\frac{1}{2} (R(\bar{Y}, \bar{Z}) X)^h - \frac{1}{4} (R(u, Y) R(u, Z) X)^h, \\ \bar{R}(X^h, Y^t) Z^h &= \frac{1}{2} (R(X, Z) \bar{Y})^t - \frac{1}{4} (R(X, R(u, Y) Z) u)^t \\ &\quad + \frac{1}{2} ((\nabla_X R)(u, Y) Z)^h, \\ \bar{R}(X^h, Y^h) Z^t &= (R(X, Y) \bar{Z})^t \\ &\quad + \frac{1}{4} (R(Y, R(u, Z) X) u - R(X, R(u, Z) Y) u)^t \\ &\quad + \frac{1}{2} ((\nabla_X R)(u, Z) Y - (\nabla_Y R)(u, Z) X)^h, \\ \bar{R}(X^h, Y^h) Z^h &= (R(X, Y) Z)^h + \frac{1}{2} (R(u, R(X, Y) u) Z)^h \\ &\quad - \frac{1}{4} (R(u, R(Y, Z) u) X - R(u, R(X, Z) u) Y)^h \\ &\quad + \frac{1}{2} ((\nabla_Z R)(X, Y) u)^t \end{aligned} \quad (2)$$

for vector fields X, Y and Z on M . (See [9].)

3. TWO TYPES OF DECOMPOSITION

Let (M^n, g) be a Riemannian manifold of dimension $n \geq 2$ and suppose that its tangent sphere bundle $T_r M$ is (locally) reducible, i.e., $(T_r M, g_S) \simeq (M_1, g_1) \times (M_2, g_2)$. A point (x, u) in $T_r M$ corresponds to a couple $(p, q) \in M_1 \times M_2$, and the tangent space $T_{(x,u)} T_r M$ can be identified with $T_p M_1 \oplus T_q M_2$. In the sequel, we will write $T_{(x,u)} M_1$ and $T_{(x,u)} M_2$ for $T_p M_1$ and $T_q M_2$, considered as subspaces of $T_{(x,u)} T_r M$, in order not to make the notation too cumbersome.

Suppose first that, at a point (x, u) of $T_r M$, the tangent space to one of the factors, say to M_1 , contains a nonzero vertical vector X^t , $X \in T_x M$ and $X \perp u$. Since we have a Riemannian product, the curvature operator $\bar{R}(\mathbf{U}, \mathbf{V})$ preserves the tangent spaces to both factors for all vectors \mathbf{U} and \mathbf{V} tangent to $T_r M$. In particular, it follows that

$$\bar{R}(Y^t, X^t)X^t = \frac{1}{r^2} (g(X, X)Y^t - g(X, Y)X^t) \in T_{(x,u)} M_1$$

for all vectors $Y \in T_x M$. As a consequence, $VT_r M_{(x,u)} \subset T_{(x,u)} M_1$, and M_1 is at least $(n-1)$ -dimensional. Hence, *if at a point of $T_r M$ one of the factors contains a nonzero vertical vector, it contains the complete vertical distribution at that point.* We call the decomposition *vertical* at (x, u) in such a situation. Note that this is the case as soon as $\max\{\dim M_1, \dim M_2\} > n$. Indeed, if $\dim M_1 > n$, then

$$\begin{aligned} \dim(VT_r M_{(x,u)} \cap T_{(x,u)} M_1) &= \dim VT_r M_{(x,u)} + \dim T_{(x,u)} M_1 \\ &\quad - \dim(VT_r M_{(x,u)} + T_{(x,u)} M_1) \\ &> (n-1) + n - (2n-1) = 0. \end{aligned}$$

So, the only possibility for the decomposition not to be vertical at (x, u) is that $\dim M_1 = n$, $\dim M_2 = n-1$ (or conversely) and neither factor is tangent to a vertical vector. We call this a *diagonal decomposition* at (x, u) .

The major part of the sequel will be devoted to the diagonal case. Using a purely infinitesimal (i.e., pointwise) approach, we show that a diagonal decomposition is only possible in one specific situation. Afterwards, we study the case of a vertical decomposition.

4. DIAGONAL DECOMPOSITION

4.1. A suitable basis. In this section, we consider a diagonal decomposition $T_r M \simeq M_1 \times M_2$ at (x, u) with $\dim M_1 = n$ and $\dim M_2 = n-1$. For dimensional reasons, we have

$$\dim(T_{(x,u)} M_1 \cap HTM_{(x,u)}) > 0.$$

Let $X_n \in T_x M$ be a unit vector such that X_n^h is tangent to M_1 at (x, u) and extend it to an orthonormal basis $\{X_1, \dots, X_n\}$ of $T_x M$. If $\pi_{*(x,u)}(T_{(x,u)} M_1) \neq T_x M$, then there must be a vertical vector tangent to M_1 at (x, u) , contrary to the hypothesis. Hence, there exist well-defined vectors Y_1, \dots, Y_{n-1} orthogonal to u for which $X_1^h + Y_1^t, \dots, X_{n-1}^h + Y_{n-1}^t$ and X_n^h are tangent to M_1 at (x, u) . Clearly, they form a basis for $T_{(x,u)} M_1$, though not in general an orthonormal one. Moreover, $\{Y_1, \dots, Y_{n-1}, u\}$ is a basis for $T_x M$ too. Otherwise, there would exist a nonzero vector $Y \in T_x M$, orthogonal to u and to Y_i , $i = 1, \dots, n-1$. But then Y^t would be orthogonal to X_n^h and to $X_i^h + Y_i^t$, $i = 1, \dots, n-1$, and hence

would belong to $(T_{(x,u)}M_1)^\perp = T_{(x,u)}M_2$, contrary to the hypothesis that M_2 has no vertical tangent vector.

Next, consider the $(n-1) \times (n-1)$ matrix $\alpha = (g_x(Y_i, Y_j))_{i,j=1,\dots,n-1}$. Since this matrix is symmetric and positive definite, it can be diagonalized by a suitable orthogonal transformation:

$$P\alpha P^t = \text{diag}(\lambda_1^2, \dots, \lambda_{n-1}^2)$$

where $P = (p_{ij}) \in O(n-1)$ and $\lambda_i > 0$ for $i = 1, \dots, n-1$. If we put

$$\tilde{X}_i = \sum_{j=1}^{n-1} p_{ij} X_j, \quad \tilde{Y}_i = \frac{1}{\lambda_i} \sum_{j=1}^{n-1} p_{ij} Y_j$$

for $i = 1, \dots, n-1$, then both $\{\tilde{X}_1, \dots, \tilde{X}_{n-1}, X_n\}$ and $\{\tilde{Y}_1, \dots, \tilde{Y}_{n-1}, u/r\}$ are orthonormal bases for $T_x M$. Further, the vectors

$$\tilde{X}_i^h + \lambda_i \tilde{Y}_i^t = \sum_{j=1}^{n-1} p_{ij} (X_j^h + Y_j^t), \quad i = 1, \dots, n-1,$$

together with X_n^h span the tangent space to M_1 at (x, u) and these vectors are pairwise orthogonal. The tangent space to M_2 at (x, u) is then spanned by the orthogonal vectors

$$\lambda_i \tilde{X}_i^h - \tilde{Y}_i^t, \quad i = 1, \dots, n-1.$$

Finally, we show that all the numbers λ_i are equal. To do this, we use that $g_S(\bar{R}(\mathbf{U}, \mathbf{V})\mathbf{W}, \mathbf{T}) = 0$ at (x, u) as soon as one of the vectors involved is tangent to M_1 and another one is tangent to M_2 . In particular, for all $i, j, k, l = 1, \dots, n-1$, it follows that

$$0 = g_S(\bar{R}(\tilde{X}_j^h + \lambda_j \tilde{Y}_j^t, \tilde{Y}_k^t)(\lambda_i \tilde{X}_i^h - \tilde{Y}_i^t), \tilde{Y}_l^t).$$

Using the expressions (2) for the curvature tensor \bar{R} of $(T_r M, g_S)$, this leads to the condition

$$0 = \lambda_i (2g(R(\tilde{X}_j, \tilde{X}_i)\tilde{Y}_k, \tilde{Y}_l) - g(R(u, \tilde{Y}_l)\tilde{X}_j, R(u, \tilde{Y}_k)\tilde{X}_i)) - \frac{4\lambda_j}{r^2} (\delta_{ik}\delta_{jl} - \delta_{ij}\delta_{kl}).$$

Switching the indices i and j , as well as k and l , we find

$$0 = \lambda_j (2g(R(\tilde{X}_i, \tilde{X}_j)\tilde{Y}_l, \tilde{Y}_k) - g(R(u, \tilde{Y}_k)\tilde{X}_i, R(u, \tilde{Y}_l)\tilde{X}_j)) - \frac{4\lambda_i}{r^2} (\delta_{jl}\delta_{ik} - \delta_{ji}\delta_{lk}).$$

Using the symmetries of the curvature tensor, it then easily follows that $\lambda_i^2 = \lambda_j^2$ or $\lambda_i = \lambda_j$.

Summarizing, we have

Lemma. *If $T_r M \simeq M_1 \times M_2$ is a diagonal decomposition at (x, u) with $\dim M_1 = n$ and $\dim M_2 = n-1$, then there exist orthonormal bases $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_{n-1}, u/r\}$ of $T_x M$ and $\lambda > 0$, such that an orthogonal basis for $T_{(x,u)}M_1$ is given by*

$$X_1^h + \lambda Y_1^t, \dots, X_{n-1}^h + \lambda Y_{n-1}^t, X_n^h$$

and an orthogonal basis for $T_{(x,u)}M_2$ is given by

$$\lambda X_1^h - Y_1^t, \dots, \lambda X_{n-1}^h - Y_{n-1}^t.$$

Remark 1. The number λ has a clear geometric meaning. Take a nonzero vertical vector \mathbf{U} at (x, u) : $\mathbf{U} = \sum_{i=1}^{n-1} \alpha_i Y_i^t$ and a nonzero vector \mathbf{V} tangent to M_2 at (x, u) : $\mathbf{V} = \sum_{i=1}^{n-1} \beta_i (\lambda X_i^h - Y_i^t)$. The angle between the two vectors has cosine given by

$$\cos(\widehat{\mathbf{UV}}) = \frac{-\sum \alpha_i \beta_i}{\sqrt{\sum \alpha_i^2} \sqrt{\sum \beta_i^2} \sqrt{1 + \lambda^2}}.$$

By the Cauchy-Schwarz inequality, we have

$$-\frac{1}{\sqrt{1 + \lambda^2}} \leq \cos(\widehat{\mathbf{UV}}) \leq \frac{1}{\sqrt{1 + \lambda^2}}$$

with equality if and only if $(\alpha_1, \dots, \alpha_{n-1})$ and $(\beta_1, \dots, \beta_{n-1})$ are proportional. We conclude that the angle θ between $VT_r M_{(x,u)}$ and $T_{(x,u)} M_2$ is such that $\cos \theta = 1/\sqrt{1 + \lambda^2}$ or $\tan \theta = \lambda$. So, λ determines the angle between $VT_r M$ and M_2 at (x, u) (and hence also between $VT_r M$ and M_1 at that point).

Remark 2. Actually, we can give a stronger formulation of the lemma. To see this, consider the mapping $\pi_1: T_{(x,u)} M_1 \rightarrow VT_r M_{(x,u)}: X^h + Y^t \mapsto Y^t$. Clearly, this mapping is linear and one-to-one on $(X_n^h)^\perp$. We restrict π_1 to $(X_n^h)^\perp \cap T_{(x,u)} M_1$ and define the linear mapping

$$A: u^\perp \rightarrow X_n^\perp: Y \mapsto \lambda \pi_{*(x,u)}(\pi_1^{-1} Y^t)$$

where, as before, $\pi: T_r M \rightarrow M$ is the natural projection map. Since

$$AY_i = \lambda \pi_{*(x,u)}(\pi_1^{-1} Y_i^t) = \lambda \pi_{*(x,u)}((X_i^h + \lambda Y_i^t)/\lambda) = X_i,$$

the map A is an isometry from u^\perp to X_n^\perp . It associates to a vector X , orthogonal to X_n , the unique vector Y , orthogonal to u , such that $X^h + \lambda Y^t$ is tangent to M_1 at (x, u) (or such that $\lambda X^h - Y^t$ is tangent to M_2 at (x, u)). So, in the lemma, we can actually choose an arbitrary orthonormal basis $\{X_1, \dots, X_{n-1}\}$ of X_n^\perp (or, alternatively, an arbitrary orthonormal basis $\{Y_1, \dots, Y_{n-1}\}$ of u^\perp). We will use this possibility in the subsequent subsections. The vectors X_n (up to sign) and u , on the other hand, are determined geometrically.

4.2. Curvature conditions. Since $(T_r M, g_S)$ is a (local) Riemannian product, all the expressions of the form $\bar{R}(\mathbf{U}, \mathbf{V})\mathbf{W}$ are zero when \mathbf{U} is tangent to M_1 and \mathbf{W} is tangent to M_2 at (x, u) . Using the curvature formulas (2), this leads to a number of curvature conditions for the manifold M . We list some of these now. From now on, indices i, j, k and l belong to $\{1, \dots, n-1\}$ unless stated otherwise.

The tangential and horizontal components of $\bar{R}(X_n^h, Y_j^t)(\lambda X_k^h - Y_k^t)$ give rise to

$$(3) \quad 2R(X_n, X_k)Y_j - \frac{2}{r^2} g(R(X_n, X_k)Y_j, u)u = R(X_n, R(u, Y_j)X_k)u,$$

$$(4) \quad 2\lambda(\nabla_{X_n} R)(u, Y_j)X_k = -2R(Y_j, Y_k)X_n - R(u, Y_j)R(u, Y_k)X_n,$$

while $\bar{R}(X_n^h, X_j^h)(\lambda X_k^h - Y_k^t) = 0$ leads to

$$(5) \quad 2\lambda(\nabla_{X_k} R)(X_n, X_j)u = 4R(X_n, X_j)Y_k - \frac{4}{r^2} g(R(X_n, X_j)Y_k, u)u \\ + R(X_j, R(u, Y_k)X_n)u - R(X_n, R(u, Y_k)X_j)u,$$

$$(6) \quad 2((\nabla_{X_n} R)(u, Y_k)X_j - (\nabla_{X_j} R)(u, Y_k)X_n) \\ = 4\lambda R(X_n, X_j)X_k + 2\lambda R(u, R(X_n, X_j)u)X_k \\ - \lambda R(u, R(X_j, X_k)u)X_n + \lambda R(u, R(X_n, X_k)u)X_j.$$

Considering $\bar{R}(X_i^h + \lambda Y_i^t, Y_j^t)(\lambda X_k^h - Y_k^t) = 0$, we obtain

$$\begin{aligned}
 (7) \quad & 2R(X_i, X_k)Y_j - \frac{2}{r^2} g(R(X_i, X_k)Y_j, u)u \\
 &= R(X_i, R(u, Y_j)X_k)u + \frac{4}{r^2} (\delta_{jk}Y_i - \delta_{ik}Y_j), \\
 (8) \quad & 2\lambda(\nabla_{X_i}R)(u, Y_j)X_k + 2R(Y_j, Y_k)X_i + R(u, Y_j)R(u, Y_k)X_i \\
 &+ \lambda^2(4R(Y_i, Y_j)X_k + R(u, Y_i)R(u, Y_j)X_k - R(u, Y_j)R(u, Y_i)X_k) = 0.
 \end{aligned}$$

Finally, from $\bar{R}(X_i^h + \lambda Y_i^t, X_j^h)(\lambda X_k^h - Y_k^t) = 0$, we derive

$$\begin{aligned}
 (9) \quad & 2\lambda(\nabla_{X_k}R)(X_i, X_j)u - 4R(X_i, X_j)Y_k + \frac{4}{r^2} g(R(X_i, X_j)Y_k, u)u \\
 &- R(X_j, R(u, Y_k)X_i)u + R(X_i, R(u, Y_k)X_j)u - 2\lambda^2 R(X_j, X_k)Y_i \\
 &+ \frac{2\lambda^2}{r^2} g(R(X_j, X_k)Y_i, u)u + \lambda^2 R(X_j, R(u, Y_i)X_k)u = 0, \\
 (10) \quad & 4\lambda R(X_i, X_j)X_k + 2\lambda R(u, R(X_i, X_j)u)X_k - \lambda R(u, R(X_j, X_k)u)X_i \\
 &+ \lambda R(u, R(X_i, X_k)u)X_j - 2(\nabla_{X_i}R)(u, Y_k)X_j + 2(\nabla_{X_j}R)(u, Y_k)X_i \\
 &- 2\lambda^2(\nabla_{X_j}R)(u, Y_i)X_k - 2\lambda R(Y_i, Y_k)X_j - \lambda R(u, Y_i)R(u, Y_k)X_j = 0.
 \end{aligned}$$

These conditions can be rewritten in an easier form. To start, we take the inner product of (3) with Y_l . This gives

$$\begin{aligned}
 2g(R(X_n, X_k)Y_j, Y_l) &= g(R(X_n, R(u, Y_j)X_k)u, Y_l) \\
 &= g(R(u, Y_l)X_n, R(u, Y_j)X_k) \\
 &= -g(R(u, Y_j)R(u, Y_l)X_n, X_k).
 \end{aligned}$$

This is equivalent to

$$(11) \quad 2R(Y_j, Y_l)X_n + R(u, Y_j)R(u, Y_l)X_n = -g(R(u, Y_j)X_n, R(u, Y_l)X_n)X_n.$$

By interchanging the indices j and l in this expression and adding both formulas, respectively subtracting them, we get

$$\begin{aligned}
 (12) \quad & R(u, Y_j)R(u, Y_l)X_n + R(u, Y_l)R(u, Y_j)X_n \\
 &= -2g(R(u, Y_j)X_n, R(u, Y_l)X_n)X_n, \\
 (13) \quad & R(u, Y_j)R(u, Y_l)X_n - R(u, Y_l)R(u, Y_j)X_n = 4R(Y_l, Y_j)X_n.
 \end{aligned}$$

Substituting (11) in (4), we find the simpler form

$$(14) \quad 2\lambda(\nabla_{X_n}R)(u, Y_j)X_k = g(R(u, Y_j)X_n, R(u, Y_k)X_n)X_n.$$

Next, we substitute (3) in (5) to obtain

$$2\lambda(\nabla_{X_k}R)(X_n, X_j)u = R(X_n, R(u, Y_k)X_j)u + R(X_j, R(u, Y_k)X_n)u.$$

Taking the inner product with Y_l , we get

$$\begin{aligned}
 & 2\lambda g((\nabla_{X_k}R)(X_n, X_j)u, Y_l) \\
 &= g(R(X_n, R(u, Y_k)X_j)u, Y_l) + g(R(X_j, R(u, Y_k)X_n)u, Y_l) \\
 &= g(R(u, Y_l)X_n, R(u, Y_k)X_j) + g(R(u, Y_l)X_j, R(u, Y_k)X_n) \\
 &= -g(R(u, Y_k)R(u, Y_l)X_n, X_j) - g(R(u, Y_l)R(u, Y_k)X_n, X_j) \\
 &= 0
 \end{aligned}$$

by (12). Hence,

$$(15) \quad (\nabla_{X_k} R)(X_n, X_j)u = 0 \quad \text{or, equivalently,} \quad (\nabla_{X_k} R)(u, Y_l)X_n = 0.$$

Substituting (14) and (15) in (6), we find

$$(16) \quad \begin{aligned} & \frac{1}{\lambda^2} g(R(u, Y_j)X_n, R(u, Y_k)X_n) X_n \\ &= 4R(X_n, X_j)X_k + 2R(u, R(X_n, X_j)u)X_k \\ & \quad - R(u, R(X_j, X_k)u)X_n + R(u, R(X_n, X_k)u)X_j. \end{aligned}$$

In order to rewrite (7), we proceed as with (3): we take the inner product with Y_l , and we use curvature properties to obtain

$$(17) \quad 2R(Y_j, Y_l)X_i + R(u, Y_j)R(u, Y_l)X_i = \frac{4}{r^2} (\delta_{il}X_j - \delta_{jl}X_i).$$

(Note that we also need (11) to know that the left-hand side in (17) is orthogonal to X_n .) Again switching the indices j and l and adding and subtracting the two formulas, we get

$$(18) \quad R(u, Y_j)R(u, Y_l)X_i + R(u, Y_l)R(u, Y_j)X_i = \frac{4}{r^2} (\delta_{il}X_j - 2\delta_{jl}X_i + \delta_{ij}X_l),$$

$$(19) \quad R(u, Y_j)R(u, Y_l)X_i - R(u, Y_l)R(u, Y_j)X_i = 4R(Y_l, Y_j)X_i + \frac{4}{r^2} (\delta_{il}X_j - \delta_{ij}X_l).$$

Substituting (17) and (19) in (8), this reduces to

$$(20) \quad \lambda(\nabla_{X_i} R)(u, Y_j)X_k = \frac{2(\lambda^2 - 1)}{r^2} (\delta_{ik}X_j - \delta_{jk}X_i),$$

or equivalently, via (15), to

$$(21) \quad \lambda(\nabla_{X_i} R)(X_k, X_l)u = \frac{2(\lambda^2 - 1)}{r^2} (\delta_{ik}Y_l - \delta_{il}Y_k).$$

It is now easily verified that (9) is a consequence of the above formulas. As to (10), using (17) and (20), it simplifies to

$$(22) \quad \begin{aligned} & 4R(X_i, X_j)X_k + 2R(u, R(X_i, X_j)u)X_k - R(u, R(X_j, X_k)u)X_i \\ & + R(u, R(X_i, X_k)u)X_j = \frac{4(\lambda^4 - \lambda^2 + 1)}{\lambda^2 r^2} (\delta_{jk}X_i - \delta_{ik}X_j). \end{aligned}$$

In the rest of this section, we will only need the formulas (12), (13), (16), (18), (19) and (22).

4.3. Clifford structures. Putting $j = l$ in (12) and (18), we see that

$$\begin{aligned} R(u, Y_j)^2 X_j &= 0, \\ R(u, Y_j)^2 X_i &= -\frac{4}{r^2} X_i, \quad i \neq j, \\ R(u, Y_j)^2 X_n &= -|R(u, Y_j)X_n|^2 X_n. \end{aligned}$$

Since $R(u, Y_j)$ is a skew-symmetric operator, the nonzero eigenvalues of $R(u, Y_j)^2$ must have even multiplicity. Hence,

- if n is even, the eigenvalue $-4/r^2$ has even multiplicity $n - 2$ on $\{X_j, X_n\}^\perp$. Hence, the eigenvalue corresponding to X_n must be zero. This implies that $R(u, Y_j)X_n = 0$ for $j = 1, \dots, n - 1$. By (13), also $R(Y_j, Y_k)X_n = 0$ for $j, k = 1, \dots, n - 1$. We conclude that X_n belongs to the nullity distribution

of the curvature tensor R_x . In this case, the conditions (12), (13) and (16) are trivially satisfied;

- if n is odd, the eigenvalue $-4/r^2$ has odd multiplicity $n - 2$ on $\{X_j, X_n\}^\perp$. So, the eigenvalue corresponding to X_n must be $-4/r^2$ as well. Hence, it follows that $|R(u, Y_j)X_n|^2 = 4/r^2$ for $j = 1, \dots, n - 1$. By Remark 2, we even have $|R(u, Y)X_n|^2 = 4/r^2$ for every unit vector Y orthogonal to u . Polarizing this identity, we obtain $g(R(u, Y)X_n, R(u, Z)X_n) = (4/r^2)g(Y, Z)$ for all vectors Y and Z orthogonal to u . In particular, the right-hand side of (12) equals $-(8\delta_{jl}/r^2)X_n$. In this case, conditions (12) and (13) are included in (18) and (19) if we allow the index i to be n .

Next, we put $i = j \neq l$ in (18). Since $R(u, Y_j)X_j = 0$ (this follows from $R(u, Y_j)^2 X_j = 0$), we obtain $R(u, Y_j)R(u, Y_l)X_j = (4/r^2)X_l$. Applying the operator $R(u, Y_j)$ on both sides, we have

$$\begin{aligned} \frac{4}{r^2} R(u, Y_j)X_l &= R(u, Y_j)^2 R(u, Y_l)X_j \\ &= -\frac{4}{r^2} (R(u, Y_l)X_j - g(R(u, Y_l)X_j, X_n)X_n) \\ &\quad - g(R(u, Y_l)X_j, X_n)|R(u, Y_j)X_n|^2 X_n \end{aligned}$$

or, equivalently,

$$4(R(u, Y_j)X_l + R(u, Y_l)X_j) = (4 - r^2|R(u, Y_j)X_n|^2)g(R(u, Y_l)X_j, X_n)X_n.$$

Since the right-hand side of this expression vanishes both when n is odd and when n is even, we conclude

$$(23) \quad R(u, Y_j)X_l + R(u, Y_l)X_j = 0$$

for $j, l = 1, \dots, n - 1$.

We are now ready to discover Clifford representations in our formulas, in particular in (12) and (18). First, consider the case when n is even. For $j = 1, \dots, n - 1$, define the operators \mathbf{R}_i acting on $V^n = T_x M$ by

$$\mathbf{R}_i = \frac{r}{2} R(u, Y_i) - \langle X_n, \cdot \rangle X_i + \langle X_i, \cdot \rangle X_n$$

where $\langle \cdot, \cdot \rangle = g_x$. In particular, it follows that $\mathbf{R}_i X_i = X_n$, $\mathbf{R}_i X_n = -X_i$ and $\mathbf{R}_i X_j = (r/2)R(u, Y_i)X_j$, $j \neq i$. Clearly, \mathbf{R}_i is a skew-symmetric operator and $\mathbf{R}_i^2 = -\text{id}$.

For $i \neq j \neq k \neq i$, we calculate:

$$\begin{aligned} (\mathbf{R}_i \circ \mathbf{R}_j + \mathbf{R}_j \circ \mathbf{R}_i) X_n &= -\mathbf{R}_i X_j - \mathbf{R}_j X_i \\ &= -\frac{r}{2} (R(u, Y_i) X_j + R(u, Y_j) X_i) = 0, \quad (\text{by (23)}) \\ (\mathbf{R}_i \circ \mathbf{R}_j + \mathbf{R}_j \circ \mathbf{R}_i) X_i &= \mathbf{R}_i \left(\frac{r}{2} R(u, Y_j) X_i \right) + \mathbf{R}_j X_n \\ &= \frac{r^2}{4} R(u, Y_i) R(u, Y_j) X_i - X_j = 0 \quad (\text{by (18)}) \\ (\mathbf{R}_i \circ \mathbf{R}_j + \mathbf{R}_j \circ \mathbf{R}_i) X_k &= \mathbf{R}_i \left(\frac{r}{2} R(u, Y_j) X_k \right) + \mathbf{R}_j \left(\frac{r}{2} R(u, Y_i) X_k \right) \\ &= \frac{r^2}{4} \left(R(u, Y_i) \{ R(u, Y_j) X_k - g(R(u, Y_j) X_k, X_i) X_i \} \right. \\ &\quad \left. + R(u, Y_j) \{ R(u, Y_i) X_k - g(R(u, Y_i) X_k, X_j) X_j \} \right) \\ &\quad + \frac{r}{2} (g(R(u, Y_j) X_k, X_i) + g(R(u, Y_i) X_k, X_j)) X_n \\ &= \frac{r^2}{4} (R(u, Y_i) R(u, Y_j) X_k + R(u, Y_j) R(u, Y_i) X_k) \\ &\quad - \frac{r}{2} g(R(u, Y_j) X_i + R(u, Y_i) X_j, X_k) X_n \\ &= 0 \quad (\text{by (18) and (23)}). \end{aligned}$$

So, for $i, j = 1, \dots, n - 1$, the operators \mathbf{R}_i satisfy

$$\mathbf{R}_i \circ \mathbf{R}_j + \mathbf{R}_j \circ \mathbf{R}_i = -2\delta_{ij} \text{ id}$$

and they correspond to a Clifford representation of an $(n - 1)$ -dimensional Clifford algebra on an n -dimensional vector space.

It is well known (see, e.g., [1] or [3]) that a given real Clifford algebra, say of dimension m , has only one (if $m \not\equiv 3 \pmod{4}$) or two (if $m \equiv 3 \pmod{4}$) irreducible representations and that the dimension n_0 of the corresponding irreducible Clifford module is completely determined by m . This relationship is given in the following table.

m	$8p$	$8p + 1$	$8p + 2$	$8p + 3$	$8p + 4$	$8p + 5$	$8p + 6$	$8p + 7$
n_0	2^{4p}	2^{4p+1}	2^{4p+2}	2^{4p+2}	2^{4p+3}	2^{4p+3}	2^{4p+3}	2^{4p+3}

For a reducible Clifford module, the dimension is a multiple kn_0 of the number n_0 corresponding to the appropriate Clifford algebra.

In the present situation, we have $m = n - 1$ and $kn_0 = n$ for even n . Therefore:

- if $n = 8p$: $8p = k2^{4p-1}$ and hence $p = 1, k = 1$ and $n = 8$;
- if $n = 8p + 2$: $8p + 2 = k2^{4p+1}$ and hence $p = 0, k = 1$ and $n = 2$;
- if $n = 8p + 4$: $8p + 4 = k2^{4p+2}$ and hence $p = 0, k = 1$ and $n = 4$;
- if $n = 8p + 6$: $8p + 6 = k2^{4p+3}$, which has no solutions.

Next, suppose that n is odd. Now, we define operators $\mathbf{R}_i, i = 1, \dots, n - 1$, acting on $V^{n+1} = T_x M \oplus \mathbb{R} X_0$ by

$$\mathbf{R}_i = \frac{r}{2} R(u, Y_i) - \langle X_0, \cdot \rangle X_i + \langle X_i, \cdot \rangle X_0$$

where $\langle \cdot, \cdot \rangle = g_x \oplus g_0$ with $g_0(aX_0, bX_0) = ab$. Precisely as before, we show that $\mathbf{R}_i \circ \mathbf{R}_j + \mathbf{R}_j \circ \mathbf{R}_i = -2\delta_{ij} \text{id}$ for $i, j = 1, \dots, n-1$. So, we have again a Clifford representation, this time with $m = n-1$ and $kn_0 = n+1$ for odd n . Therefore, by the table above:

- if $n = 8p + 1$: $8p + 2 = k2^{4p}$ and hence $p = 0$, $k = 2$ and $n = 1$;
- if $n = 8p + 3$: $8p + 4 = k2^{4p+2}$ and hence $p = 0$, $k = 1$ and $n = 3$;
- if $n = 8p + 5$: $8p + 6 = k2^{4p+3}$, which has no solutions;
- if $n = 8p + 7$: $8p + 8 = k2^{4p+3}$ and hence $p = 0$, $k = 1$ and $n = 7$.

We conclude from this subsection that diagonal decompositions can only occur when the base manifold has dimension 2, 3, 4, 7 or 8. (The case $n = 1$ is irrelevant, since then $T_r M$ has dimension equal to one and no decompositions exist.)

4.4. The remaining dimensions.

Case $n = 2$. In this situation, we have a two-dimensional manifold for which the nullity vector space of the curvature tensor is non-trivial. This implies that the curvature tensor is identically zero and the space is flat.

Conversely, since any tangent sphere bundle of a flat surface $M^2(0)$ is a flat three-dimensional space, a diagonal decomposition actually exists around each point (x, u) of $T_r M^2(0)$. Note, however, that we also have $T_r M^2(0) \simeq M^2(0) \times S^1(r)$ with $\{x\} \times S^1(r) \simeq \pi^{-1}(x)$. So, $T_r M^2(0)$ also admits a vertical decomposition.

Case $n = 3$. Let X_3 be the unique unit vector (up to sign) such that X_3^h is tangent to M_1 at (x, u) . Pick a unit vector X_1 orthogonal to X_3 and let Y_1 be the corresponding unit vector orthogonal to u (i.e., $X_1^h + \lambda Y_1^t$ is tangent to M_1). From the comments at the beginning of Subsection 4.3, we know that $(r/2)R(u, Y_1)X_3$ is a unit vector, which is moreover orthogonal to X_1 and X_3 . So, we obtain an orthonormal basis $\{X_1, X_2, X_3\}$ by defining X_2 to be $X_2 := (r/2)R(u, Y_1)X_3$. Let Y_2 be the corresponding unit vector orthogonal to u and Y_1 . (Since each Y_i is fixed together with its corresponding X_i , we will not mention this explicitly anymore in what follows.)

Using the properties of the operators $R(u, Y_1)$ and $R(u, Y_2)$, we then deduce that

$$(24) \quad \begin{aligned} R(u, Y_1)X_1 &= 0, & R(u, Y_1)X_2 &= -\frac{2}{r}X_3, & R(u, Y_1)X_3 &= \frac{2}{r}X_2, \\ R(u, Y_2)X_1 &= \frac{2}{r}X_3, & R(u, Y_2)X_2 &= 0, & R(u, Y_2)X_3 &= -\frac{2}{r}X_1 \end{aligned}$$

and from (13) and (19) it follows that

$$(25) \quad R(Y_1, Y_2)X_1 = -\frac{2}{r^2}X_2, \quad R(Y_1, Y_2)X_2 = \frac{2}{r^2}X_1, \quad R(Y_1, Y_2)X_3 = 0.$$

Next, we compute $R(X_i, X_j)X_k$, $i, j, k = 1, 2, 3$, from the equalities (16) and (22), writing $R(u, R(X_i, X_j)u)X_k$ as $\sum g(R(u, Y_l)X_i, X_j)R(u, Y_l)X_k$ and using (24) and (25). This gives

$$(26) \quad \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline r^2 R(X_1, X_2) & -AX_2 & AX_1 & 0 \\ r^2 R(X_1, X_3) & -CX_3 & 0 & CX_1 \\ r^2 R(X_2, X_3) & 0 & -CX_3 & CX_3 \end{array}$$

where $A = (\lambda^4 - \lambda^2 + 1)/\lambda^2$ and $C = (3\lambda^2 + 1)/\lambda^2$.

Since both $\{X_1, X_2, X_3\}$ and $\{Y_1, Y_2, u/r\}$ are orthonormal bases for $T_x M$, there is an orthogonal matrix $Q = (q_{ij}) \in O(3)$ such that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = Q \begin{pmatrix} Y_1 \\ Y_2 \\ u/r \end{pmatrix}.$$

Changing X_3 to $-X_3$ if necessary, we may even suppose that $Q \in SO(3)$. Then

$$\begin{aligned} R(X_1, X_3) &= (q_{11}q_{32} - q_{12}q_{31}) R(Y_1, Y_2) + \frac{q_{11}q_{33} - q_{13}q_{31}}{r} R(Y_1, u) \\ &\quad + \frac{q_{12}q_{33} - q_{13}q_{32}}{r} R(Y_2, u) \\ &= -q_{23} R(Y_1, Y_2) - \frac{q_{22}}{r} R(u, Y_1) + \frac{q_{21}}{r} R(u, Y_2). \end{aligned}$$

If we let both sides act on X_1, X_2 and X_3 and if we use (24), (25) and (26), we find that

$$q_{21} = -C/2, \quad q_{22} = 0, \quad q_{23} = 0.$$

Since $Q \in SO(3)$, it follows that $q_{21}^2 + q_{22}^2 + q_{23}^2 = 1$ and hence $1 = (3\lambda^2 + 1)/2\lambda^2$ or $\lambda^2 + 1 = 0$, which is a contradiction. Hence, no three-dimensional manifold admits a diagonal decomposition of its tangent sphere bundles at any point.

Case $n = 4$. Let X_4 be the unique unit vector (up to sign) in the nullity distribution of R_x . Take two mutually orthogonal unit vectors X_1 and X_2 perpendicular to X_4 . Since $(r/2)R(u, Y_1)X_2$ is a unit vector and orthogonal to X_1, X_2 and X_4 , we can define $X_3 := (r/2)R(u, Y_1)X_2$. From the properties of the operators $R(u, Y_i)$, $i = 1, 2, 3$, it follows that

	X_1	X_2	X_3	X_4
(27) $r R(u, Y_1)$	0	$2X_3$	$-2X_2$	0
$r R(u, Y_2)$	$-2X_3$	0	$2X_1$	0
$r R(u, Y_3)$	$2X_2$	$-2X_1$	0	0

Next, we decompose X_4 with respect to the basis $\{Y_1, Y_2, Y_3, u/r\}$:

$$X_4 = q_1 Y_1 + q_2 Y_2 + q_3 Y_3 + q_4 \frac{u}{r}, \quad q_1^2 + q_2^2 + q_3^2 + q_4^2 = 1.$$

Then $R(u, X_4) = q_1 R(u, Y_1) + q_2 R(u, Y_2) + q_3 R(u, Y_3)$. Since X_4 belongs to the nullity distribution of R , this operator vanishes identically. By (27), we must have $q_1 = q_2 = q_3 = 0$. Hence, $X_4 = \pm u/r$. But this is impossible, since u clearly does not belong to the nullity distribution. So, also for four-dimensional manifolds, a diagonal decomposition of its tangent sphere bundles does not exist at any point.

Case $n = 7$. The argument for $n = 7$ goes along the same lines as that for $n = 3$, but it is more involved technically. Again we start with the unit vector X_7 , uniquely

determined up to sign, such that X_7^h is tangent to M_1 , and with an arbitrary unit vector X_1 orthogonal to X_7 . The unit vector $X_2 := (r/2)R(u, Y_1)X_7$ is orthogonal to both X_1 and X_7 . Then it follows that

$$\begin{aligned} R(u, Y_1)X_1 &= 0, & R(u, Y_1)X_2 &= -\frac{2}{r}X_7, & R(u, Y_1)X_7 &= \frac{2}{r}X_2, \\ R(u, Y_2)X_1 &= \frac{2}{r}X_7, & R(u, Y_2)X_2 &= 0, & R(u, Y_2)X_7 &= -\frac{2}{r}X_1. \end{aligned}$$

Note that $R(u, Y_1)$ and $R(u, Y_2)$ preserve $\text{span}\{X_1, X_2, X_7\}$, hence by skew-symmetry also its orthogonal complement. Next, take a unit vector X_4 orthogonal to X_1, X_2, X_7 and define the unit vectors $X_5 := (r/2)R(u, Y_2)X_4$ and $X_6 := (r/2)R(u, Y_1)X_4$. Then X_5 and X_6 are already orthogonal to X_1, X_2, X_4 and X_7 . Further,

$$\begin{aligned} g(X_5, X_6) &= \frac{r^2}{4} g(R(u, Y_2)X_4, R(u, Y_1)X_4) \\ &= -\frac{r^2}{4} g(R(u, Y_1)R(u, Y_2)X_4, X_4) \\ (28) \quad &= \frac{r^2}{4} g(R(u, Y_2)R(u, Y_1)X_4, X_4) \quad (\text{by (18)}) \\ &= -\frac{r^2}{4} g(R(u, Y_1)X_4, R(u, Y_2)X_4) \\ &= -g(X_5, X_6) \end{aligned}$$

and X_5 and X_6 are mutually orthogonal as well. Finally, since $R(u, Y_1)X_5$ is orthogonal to X_1, X_2, X_5, X_7 and

$$\begin{aligned} g(R(u, Y_1)X_5, X_4) &= -g(X_5, R(u, Y_1)X_4) = -\frac{2}{r}g(X_5, X_6) = 0, \\ g(R(u, Y_1)X_5, X_6) &= \frac{r}{2}g(R(u, Y_1)X_5, R(u, Y_1)X_4) = \frac{2}{r}g(X_5, X_4) = 0, \end{aligned}$$

we may define $X_3 := (r/2)R(u, Y_1)X_5$.

In this way, we have defined an orthonormal basis $\{X_1, \dots, X_7\}$, and the actions of the operators $R(u, Y_i)$, $i = 1, \dots, 6$, can be computed explicitly in this basis using the properties (12), (18) and (23) above. We obtain

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$r R(u, Y_1)$	0	$-2X_7$	$-2X_5$	$2X_6$	$2X_3$	$-2X_4$	$2X_2$
$r R(u, Y_2)$	$2X_7$	0	$2X_6$	$2X_5$	$-2X_4$	$-2X_3$	$-2X_1$
$r R(u, Y_3)$	$2X_5$	$-2X_6$	0	$-2X_7$	$-2X_1$	$2X_2$	$2X_4$
$r R(u, Y_4)$	$-2X_6$	$-2X_5$	$2X_7$	0	$2X_2$	$2X_1$	$-2X_3$
$r R(u, Y_5)$	$-2X_3$	$2X_4$	$2X_1$	$-2X_2$	0	$-2X_7$	$2X_6$
$r R(u, Y_6)$	$2X_4$	$2X_3$	$-2X_2$	$-2X_1$	$2X_7$	0	$-2X_5$

Next, we calculate $R(Y_i, Y_j)X_k$ from (13) and (19):

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$r^2 R(Y_1, Y_2)$	$-2X_2$	$2X_1$	$2X_4$	$-2X_3$	$2X_6$	$-2X_5$	0
$r^2 R(Y_1, Y_3)$	$-2X_3$	$-2X_4$	$2X_1$	$2X_2$	0	$2X_7$	$-2X_6$
$r^2 R(Y_1, Y_4)$	$-2X_4$	$2X_3$	$-2X_2$	$2X_1$	$2X_7$	0	$-2X_5$
$r^2 R(Y_1, Y_5)$	$-2X_5$	$-2X_6$	0	$-2X_7$	$2X_1$	$2X_2$	$2X_4$
$r^2 R(Y_1, Y_6)$	$-2X_6$	$2X_5$	$-2X_7$	0	$-2X_2$	$2X_1$	$2X_3$
$r^2 R(Y_2, Y_3)$	$2X_4$	$-2X_3$	$2X_2$	$-2X_1$	$2X_7$	0	$-2X_5$
$r^2 R(Y_2, Y_4)$	$-2X_3$	$-2X_4$	$2X_1$	$2X_2$	0	$-2X_7$	$2X_6$
$r^2 R(Y_2, Y_5)$	$2X_6$	$-2X_5$	$-2X_7$	0	$2X_2$	$-2X_1$	$2X_3$
$r^2 R(Y_2, Y_6)$	$-2X_5$	$-2X_6$	0	$2X_7$	$2X_1$	$2X_2$	$-2X_4$
$r^2 R(Y_3, Y_4)$	$2X_2$	$-2X_1$	$-2X_4$	$2X_3$	$2X_6$	$-2X_5$	0
$r^2 R(Y_3, Y_5)$	0	$2X_7$	$-2X_5$	$-2X_6$	$2X_3$	$2X_4$	$-2X_2$
$r^2 R(Y_3, Y_6)$	$2X_7$	0	$-2X_6$	$2X_5$	$-2X_4$	$2X_3$	$-2X_1$
$r^2 R(Y_4, Y_5)$	$2X_7$	0	$2X_6$	$-2X_5$	$2X_4$	$-2X_3$	$-2X_1$
$r^2 R(Y_4, Y_6)$	0	$-2X_7$	$-2X_5$	$-2X_6$	$2X_3$	$2X_4$	$2X_2$
$r^2 R(Y_5, Y_6)$	$2X_2$	$-2X_1$	$2X_4$	$-2X_3$	$-2X_6$	$2X_5$	0

Using (16) and (22), we can now compute the curvature components $R(X_i, X_j)X_k$ for $i, j, k = 1, \dots, 7$:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$r^2 R(X_1, X_2)$	$-AX_2$	AX_1	$2X_4$	$-2X_3$	$2X_6$	$-2X_5$	0
$r^2 R(X_1, X_3)$	$-BX_3$	X_4	BX_1	$-X_2$	0	0	0
$r^2 R(X_1, X_4)$	$-BX_4$	$-X_3$	X_2	BX_1	0	0	0
$r^2 R(X_1, X_5)$	$-BX_5$	X_6	0	0	BX_1	$-X_2$	0
$r^2 R(X_1, X_6)$	$-BX_6$	$-X_5$	0	0	X_2	BX_1	0
$r^2 R(X_1, X_7)$	$-CX_7$	0	0	0	0	0	CX_1
$r^2 R(X_2, X_3)$	$-X_4$	$-BX_3$	BX_2	X_1	0	0	0
$r^2 R(X_2, X_4)$	X_3	$-BX_4$	$-X_1$	BX_2	0	0	0
$r^2 R(X_2, X_5)$	$-X_6$	$-BX_5$	0	0	BX_2	X_1	0
$r^2 R(X_2, X_6)$	X_5	$-BX_6$	0	0	$-X_1$	BX_2	0
$r^2 R(X_2, X_7)$	0	$-CX_7$	0	0	0	0	CX_2
$r^2 R(X_3, X_4)$	$2X_2$	$-2X_1$	$-AX_4$	AX_3	$2X_6$	$-2X_5$	0
$r^2 R(X_3, X_5)$	0	0	$-BX_5$	X_6	BX_3	$-X_4$	0
$r^2 R(X_3, X_6)$	0	0	$-BX_6$	$-X_5$	X_4	BX_3	0
$r^2 R(X_3, X_7)$	0	0	$-CX_7$	0	0	0	CX_3
$r^2 R(X_4, X_5)$	0	0	$-X_6$	$-BX_5$	BX_4	X_3	0
$r^2 R(X_4, X_6)$	0	0	X_5	$-BX_6$	$-X_3$	BX_4	0
$r^2 R(X_4, X_7)$	0	0	0	$-CX_7$	0	0	CX_4
$r^2 R(X_5, X_6)$	$2X_2$	$-2X_1$	$2X_4$	$-2X_3$	$-AX_6$	AX_5	0
$r^2 R(X_5, X_7)$	0	0	0	0	$-CX_7$	0	CX_5
$r^2 R(X_6, X_7)$	0	0	0	0	0	$-CX_7$	CX_6

where $A = (\lambda^4 - \lambda^2 + 1)\lambda^2$, $B = (\lambda^2 + 1)^2/\lambda^2$ and $C = (3\lambda^2 + 1)/\lambda^2$.

We now show that the tables above are incompatible. To see this, we relate the two orthonormal bases $\{X_1, \dots, X_7\}$ and $\{Y_1, \dots, Y_6, u/r\}$ by an orthogonal

transformation. Let $Q = (q_{ij}) \in O(7)$ be such that

$$\begin{pmatrix} X_1 \\ \vdots \\ X_7 \end{pmatrix} = Q \begin{pmatrix} Y_1 \\ \vdots \\ Y_6 \\ u/r \end{pmatrix}.$$

Putting $Q_{kl}^{ij} := q_{ik}q_{jl} - q_{il}q_{jk}$, we then have the equality

$$R(X_i, X_j) = \sum_{k < l=1}^6 Q_{kl}^{ij} R(Y_k, Y_l) + \sum_{k=1}^6 (Q_{k7}^{ij}/r) R(Y_k, u).$$

So,

$$\begin{aligned} 2 &= r^2 g(R(X_1, X_2)X_3, X_4) \\ &= \sum_{k < l=1}^6 r^2 Q_{kl}^{12} g(R(Y_k, Y_l)X_3, X_4) + \sum_{k=1}^6 r Q_{k7}^{12} g(R(Y_k, u)X_3, X_4) \\ &= 2(Q_{12}^{12} - Q_{34}^{12} + Q_{56}^{12}) \end{aligned}$$

and

$$2 = r^2 g(R(X_1, X_2)X_5, X_6) = 2(Q_{12}^{12} + Q_{34}^{12} - Q_{56}^{12}).$$

This implies that $Q_{12}^{12} = 1$. Now, using the Cauchy-Schwarz inequality and the fact that Q is orthogonal, we find that

$$\begin{aligned} 1 &= Q_{12}^{12} = q_{11}q_{22} - q_{12}q_{21} = (q_{11}, q_{12}) \cdot (q_{22}, -q_{21}) \\ &\leq \sqrt{q_{11}^2 + q_{12}^2} \sqrt{q_{21}^2 + q_{22}^2} \leq \sqrt{q_{11}^2 + \cdots + q_{17}^2} \sqrt{q_{21}^2 + \cdots + q_{27}^2} = 1. \end{aligned}$$

Hence, all the inequalities must be equalities. In particular, we have $q_{13} = \cdots = q_{17} = q_{23} = \cdots = q_{27} = 0$ and consequently

$$X_1 = \cos \theta_1 Y_1 + \sin \theta_1 Y_2, \quad X_2 = \epsilon_1 (-\sin \theta_1 Y_1 + \cos \theta_1 Y_2)$$

where $\epsilon_1 = \pm 1$ and θ_1 is some real number. In a similar way, we can show that $Q_{34}^{34} = Q_{56}^{56} = 1$ and that

$$\begin{aligned} X_3 &= \cos \theta_2 Y_3 + \sin \theta_2 Y_4, & X_4 &= \epsilon_2 (-\sin \theta_2 Y_3 + \cos \theta_2 Y_4), \\ X_5 &= \cos \theta_3 Y_5 + \sin \theta_3 Y_6, & X_6 &= \epsilon_3 (-\sin \theta_3 Y_5 + \cos \theta_3 Y_6). \end{aligned}$$

As a consequence, we also have $X_7 = \epsilon u/r$, $\epsilon = \pm 1$. Using the tables above, we find that

$$\begin{aligned} 0 &= r^2 R(X_1, X_7)X_3 = -\epsilon (\cos \theta_1 r R(u, Y_1)X_3 + \sin \theta_1 r R(u, Y_2)X_3) \\ &= 2\epsilon (\cos \theta_1 X_5 - \sin \theta_1 X_6), \end{aligned}$$

which gives a contradiction. So, also seven-dimensional manifolds cannot have a diagonal decomposition for their tangent sphere bundles at any point.

Case $n = 8$. This case is treated as the case $n = 4$, but the appropriate choice for the basis $\{X_1, \dots, X_8\}$ requires a little more care. Let X_8 be the unique unit vector (up to sign) in the nullity distribution of R_x and take two arbitrary unit vectors X_1 and X_2 that are mutually orthogonal and perpendicular to X_8 . As before, we

define $X_3 := (r/2)R(u, Y_1)X_2$, which is a unit vector orthogonal to X_1, X_2 and X_8 . It follows that $R(u, Y_i)X_8 = 0$ for $i = 1, 2, 3$, and

$$\begin{aligned} R(u, Y_1)X_1 &= 0, & R(u, Y_1)X_2 &= \tfrac{2}{r}X_3, & R(u, Y_1)X_3 &= -\tfrac{2}{r}X_2, \\ R(u, Y_2)X_1 &= -\tfrac{2}{r}X_3, & R(u, Y_2)X_2 &= 0, & R(u, Y_2)X_3 &= \tfrac{2}{r}X_1, \\ R(u, Y_3)X_1 &= \tfrac{2}{r}X_2, & R(u, Y_3)X_2 &= -\tfrac{2}{r}X_1, & R(u, Y_3)X_3 &= 0. \end{aligned}$$

Because they are skew-symmetric, the operators $R(u, Y_1), R(u, Y_2)$ and $R(u, Y_3)$ also preserve $W = \{X_1, X_2, X_3, X_8\}^\perp$, and on this space they anti-commute by (18). It is easy to check that the operator $(r^3/8)R(u, Y_1)R(u, Y_2)R(u, Y_3)$ is symmetric on W and that it squares to the identity on W . Hence, it has a basis of eigenvectors corresponding to the eigenvalues $+1$ and -1 . Let X_4 be a unit vector in W such that $r^3 R(u, Y_1)R(u, Y_2)R(u, Y_3)X_4 = 8\epsilon X_4$ where $\epsilon = \pm 1$, and define

$$X_5 := \tfrac{r}{2}R(u, Y_1)X_4, \quad X_6 := \tfrac{r}{2}R(u, Y_2)X_4, \quad X_7 := \tfrac{r}{2}R(u, Y_3)X_4.$$

Clearly, X_5, X_6 and X_7 are unit vectors orthogonal to X_1, X_2, X_3, X_4 and X_8 . A computation similar to (28) shows that they are also orthogonal to one another. So, we have an orthonormal basis $\{X_1, \dots, X_8\}$ for T_xM . It is now possible to compute explicitly the action of $R(u, Y_i), i = 1, \dots, 7$, from the condition (18). We get

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
$r R(u, Y_1)$	0	$2X_3$	$-2X_2$	$2X_5$	$-2X_4$	$-2\epsilon X_7$	$2\epsilon X_6$	0
$r R(u, Y_2)$	$-2X_3$	0	$2X_1$	$2X_6$	$2\epsilon X_7$	$-2X_4$	$-2\epsilon X_5$	0
$r R(u, Y_3)$	$2X_2$	$-2X_1$	0	$2X_7$	$-2\epsilon X_6$	$2\epsilon X_5$	$-2X_4$	0
$r R(u, Y_4)$	$-2X_5$	$-2X_6$	$-2X_7$	0	$2X_1$	$2X_2$	$2X_3$	0
$r R(u, Y_5)$	$2X_4$	$-2\epsilon X_7$	$2\epsilon X_6$	$-2X_1$	0	$-2\epsilon X_3$	$2\epsilon X_2$	0
$r R(u, Y_6)$	$2\epsilon X_7$	$2X_4$	$-2\epsilon X_5$	$-2X_2$	$2\epsilon X_3$	0	$-2\epsilon X_1$	0
$r R(u, Y_7)$	$-2\epsilon X_6$	$2\epsilon X_5$	$2X_4$	$-2X_3$	$-2\epsilon X_2$	$2\epsilon X_1$	0	0

Next, decompose X_8 with respect to the basis $\{Y_1, \dots, Y_7, u/r\}$:

$$X_8 = q_1 Y_1 + \dots + q_7 Y_7 + q_8 \frac{u}{r}, \quad q_1^2 + \dots + q_8^2 = 1.$$

Since X_8 belongs to the nullity distribution of the curvature, we have

$$0 = R(u, X_8) = \sum_{i=1}^7 q_i R(u, Y_i)$$

and from the table above we deduce $q_1 = \dots = q_7 = 0$. Hence, $X_8 = \pm u/r$, but this is impossible since u does not belong to the nullity distribution. So, also in the eight-dimensional case, a diagonal decomposition of the tangent sphere bundles does not exist at even a single point.

Remark 3. The operator $r^3 R(u, Y_1)R(u, Y_2)R(u, Y_3)$ acts as $8\epsilon \text{id}$ on the vector space spanned by X_4, X_5, X_6 and X_7 , as is seen easily from the previous table. The two different cases, $\epsilon = +1$ and $\epsilon = -1$, correspond to the two non-equivalent irreducible Clifford representations of the seven-dimensional Clifford algebra.

5. VERTICAL DECOMPOSITION

Now, we suppose that we have a vertical decomposition $T_r M \simeq M_1 \times M_2$ such that $VT_r M_{(x,u)} \subset T_{(x,u)} M_1$ everywhere. In this situation, if $(x, u) \in M_1 \times \{q\}$ for some $q \in M_2$, then $\pi^{-1}(x) \subset M_1 \times \{q\}$. Consequently, we have $M_1 \times \{q\} = \pi^{-1}(\pi(M_1 \times \{q\}))$. So, the leaves $M_1 \times \{q\}$, corresponding to the product, project under π to a foliation \mathbf{L}_1 on (M, g) and $\pi^{-1}(\mathbf{L}_1) = \{M_1 \times \{q\}, q \in M_2\}$. Let L_1 be the distribution on M tangent to \mathbf{L}_1 . Define the distribution L_2 to be the orthogonal distribution to L_1 on M . Then

$$T_{(x,u)}(M_1 \times \{q\}) = VT_r M_{(x,u)} \oplus h(L_{1x}), \quad T_{(x,u)}(\{p\} \times M_2) = h(L_{2x})$$

where h denotes the horizontal lift.

If X and Y are vector fields on M tangent to L_1 and U, V are tangent to L_2 , then X^h, Y^h are tangent to M_1 and U^h, V^h are tangent to M_2 . Because of the product structure, we have that $\bar{\nabla}_{X^h} Y^h$ and $\bar{\nabla}_{U^h} X^h$ are tangent to M_1 and $\bar{\nabla}_{U^h} V^h$ and $\bar{\nabla}_{X^h} U^h$ are tangent to M_2 . Using the expressions (1) for $\bar{\nabla}$, this means that

- $\nabla_X Y$ and $\nabla_U X$ are sections of L_1 : so, L_1 is totally geodesic and even totally parallel;
- $\nabla_U V$ and $\nabla_X U$ are sections of L_2 : so, also L_2 is totally geodesic and totally parallel (in particular, L_2 is integrable with associated foliation \mathbf{L}_2);
- $R(U, V)u = R(X, U)u = 0$: so, L_2 is contained in the nullity distribution of the curvature. The leaves of \mathbf{L}_2 are therefore flat.

These properties imply that \mathbf{L}_1 and \mathbf{L}_2 consist of the leaves of a local Riemannian product $M \simeq M' \times \mathbb{R}^k$ where $k = \dim L_2 \leq n$ (see [8]).

Suppose conversely that M^n is locally isometric to $M' \times \mathbb{R}^k$ for $1 \leq k \leq n$. This gives rise to two foliations on M^n : $\mathbf{L}_1 = \{M' \times \{v\}, v \in \mathbb{R}^k\}$ and $\mathbf{L}_2 = \{\{p\} \times \mathbb{R}^k, p \in M'\}$. Define two complementary distributions \tilde{L}_1 and \tilde{L}_2 on $T_r M$ by

$$\tilde{L}_1 = VT_r M \oplus h(TM'), \quad \tilde{L}_2 = h(T\mathbb{R}^k).$$

It is easily checked using (1) that \tilde{L}_1 and \tilde{L}_2 are totally geodesic and totally parallel complementary distributions. Hence, the leaves of their corresponding foliations $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{L}}_2$ are actually the leaves of a local Riemannian product. In particular, note that $\tilde{\mathbf{L}}_1 = \{\pi^{-1}(M' \times \{v\}), v \in \mathbb{R}^k\}$. So, $T_r M$ is indeed locally reducible.

6. GLOBAL RESULTS

We continue with the notation of the previous section. In order to derive results concerning the global reducibility of $(T_r M, g_S)$, we will exploit the relationship between the foliations \mathbf{L}_1 and \mathbf{L}_2 of (M, g) and the foliations $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{L}}_2$ of $(T_r M, g_S)$ in the case of a vertical decomposition. We have already remarked that \mathbf{L}_1 and $\tilde{\mathbf{L}}_1$ determine each other reciprocally by $\mathbf{L}_1 = \pi(\tilde{\mathbf{L}}_1)$ and $\tilde{\mathbf{L}}_1 = \pi^{-1}\mathbf{L}_1$. The relationship between the foliations \mathbf{L}_2 and $\tilde{\mathbf{L}}_2$ is not so straightforward. We still have $\mathbf{L}_2 = \pi(\tilde{\mathbf{L}}_2)$, but determining $\tilde{\mathbf{L}}_2$ from \mathbf{L}_2 requires a little more care. To construct the leaf \tilde{S} of $\tilde{\mathbf{L}}_2$ through a point $(x, u) \in T_r M$, consider all the curves in the leaf S of \mathbf{L}_2 starting at $x \in M$. Then, \tilde{S} consists of all end-points of the horizontal lifts of these curves starting at (x, u) . We call \tilde{S} the *horizontal lift of S through (x, u)* . Since \tilde{S} is everywhere horizontal, the map $\pi: \tilde{S} \rightarrow S$ is a local isometry and \tilde{S} is a Riemannian covering of S . When S is simply connected, \tilde{S} and S are globally isometric and, in particular, one-to-one.

With these comments in mind, we now proceed to the proof of the Global Theorem. So, we suppose that $\dim M \geq 3$ and that $(T_r M, g_S)$ is isometric to a global Riemannian product $(M_1, g_1) \times (M_2, g_2)$. Since $\dim M \geq 3$, this is a vertical decomposition and $VT_r M$ is tangent to one of the factors, say M_1 . Consider M_1 and M_2 as submanifolds of $T_r M$ (i.e., choose one leaf of each of the foliations $\tilde{\mathbf{L}}_1$ and $\tilde{\mathbf{L}}_2$) and define the submanifolds $M' := \pi(M_1)$ and $F := \pi(M_2)$ of M . From the local considerations in the previous section, we know that (M, g) is locally isometric to the Riemannian product $M' \times F$ and that F is flat.

We show that there is a one-to-one correspondence between M and $M' \times F$. Take a point $x \in M$ and consider an arbitrary vector $u \in T_x M$ of length r . Through $(x, u) \in T_r M$, there is a unique leaf \tilde{S}_{1u} of $\tilde{\mathbf{L}}_1$ and a unique leaf \tilde{S}_{2u} of $\tilde{\mathbf{L}}_2$. Because of the product structure on $T_r M$, \tilde{S}_{1u} cuts M_2 in a unique point $\tilde{q}_u \in M_2$ and \tilde{S}_{2u} cuts M_1 in a unique point $\tilde{p}_u \in M_1$. Put $p_u := \pi(\tilde{p}_u) \in M'$ and $q_u := \pi(\tilde{q}_u) \in F$. We claim that the correspondence $M \rightarrow M' \times F: x \mapsto (p_u, q_u)$ is well-defined, i.e., independent of the choice of the tangent vector u . To see this, take another vector $v \in T_x M$ of length r . Since $\pi(x, u) = x = \pi(x, v)$, the leaf \tilde{S}_{1u} of $\tilde{\mathbf{L}}_1$ contains both (x, u) and (x, v) ; so we have $\tilde{S}_{1u} = \tilde{S}_{1v}$, $\tilde{q}_u = \tilde{q}_v$ and $q_u = q_v$. The unique leaf \tilde{S}_{2v} of $\tilde{\mathbf{L}}_2$ through (x, v) is different from \tilde{S}_{2u} . However, both are horizontal lifts of $S_{2x} = \pi(\tilde{S}_{2u})$. So, if $\tilde{\gamma}_u(t) = (x(t), u(t))$ is a curve in \tilde{S}_{2u} such that $\tilde{\gamma}_u(0) = (x, u)$ and $\tilde{\gamma}_u(1) = \tilde{p}_u \in M_1$, then $\gamma = \pi \circ \tilde{\gamma}_u$ runs from x to $p_u \in M$ in S_{2x} . Denote by $\tilde{\gamma}_v$ the horizontal lift of γ starting at (x, v) . Clearly, $\tilde{\gamma}_v$ lies in \tilde{S}_{2v} and ends at $\tilde{p}_v \in M_1$. Hence, $p_v = \pi(\tilde{p}_v) = \pi(\tilde{\gamma}_v(1)) = \gamma(1) = p_u$.

On the other hand, starting from a couple $(p, q) \in M' \times F$, we find the corresponding point $x \in M$ as $x = \pi(\tilde{p}, \tilde{q})$ for some $\tilde{p} \in M_1$ with $\pi(\tilde{p}) = p$ and the unique $\tilde{q} \in M_2$ with $\pi(\tilde{q}) = q$. Via an argument as above, one shows that x does not depend on the choice of \tilde{p} and that the map $(p, q) \mapsto x$ defined in this way is the inverse of the map $x \mapsto (p_u, q_u)$.

Next, we note that the correspondence $M \rightarrow M' \times F: x \mapsto (p_u, q_u)$ is defined so as to respect the local product structure. In particular, the metric g on M corresponds to the product metric of $M' \times F$, and the first statement is proved.

Conversely, suppose that (M, g) is the global product space $(M', g') \times (F, g_0)$. By choosing a leaf of both product foliations, one can consider M' and F as submanifolds of M . Let x_0 be their intersection point and choose a vector $u_0 \in T_{x_0} M$ of length r . Define M_1 as the inverse image of M' under the projection π and M_2 as the horizontal lift of F through (x_0, u_0) . Since we suppose F to be simply connected, M_2 is isometric to the flat space (F, g_0) and M_1 and M_2 have (x_0, u_0) as unique intersection point.

We show that there is a one-to-one correspondence between $T_r M$ and $M_1 \times M_2$. Take $(x, u) \in T_r M$ and denote by S_1 the unique leaf of \mathbf{L}_1 on M through x and by S_2 the unique leaf of \mathbf{L}_2 on M through x . Then, the leaf \tilde{S}_1 of $\tilde{\mathbf{L}}_1$ through (x, u) is given by $\pi^{-1}(S_1)$ and the leaf \tilde{S}_2 of $\tilde{\mathbf{L}}_2$ through (x, u) is the horizontal lift of S_2 through this point. \tilde{S}_1 cuts M_2 in a unique point \tilde{q} with $\pi(\tilde{q}) = S_1 \cap F$, and \tilde{S}_2 cuts M_1 in a unique point \tilde{p} with $\pi(\tilde{p}) = S_2 \cap M'$. (Note that the simply connectedness of F is essential to ensure uniqueness.) Clearly, the correspondence $T_r M \rightarrow M_1 \times M_2: (x, u) \mapsto (\tilde{p}, \tilde{q})$ is well-defined and it is not difficult to construct its inverse. Since this correspondence also respects the local product structure, the metric g_S on $T_r M$ corresponds to the product metric on $M_1 \times M_2$. This completes the proof of the Global Theorem.

Remark 4. The proof of the Global Theorem continues to hold when $n = 2$ for the case of a vertical global decomposition of $(T_r M, g_S)$. Clearly, the base manifold is then flat. That we need the simply connectedness of the flat factor can be seen from the example of a two-dimensional flat cone C . The vertical and horizontal distributions on $T_r C$ are both integrable, and locally their integral manifolds are the leaves of the local product foliation on $T_r C$. If it were a global decomposition, every maximal integral manifold of the horizontal distribution would intersect every vertical fiber exactly once and it would be isometric to C under the natural projection π . This would define a global parallelization of C , contrary to the fact that its full holonomy group is non-trivial.

REFERENCES

- [1] M. F. Atiyah, R. Bott and A. Shapiro, *Clifford modules*, Topology **3**, Suppl. 1 (1964), 3–38. MR **29**:5250
- [2] J. Berndt, E. Boeckx, P. Nagy and L. Vanhecke, *Geodesics on the unit tangent bundle*, preprint, 2001.
- [3] J. Berndt, F. Tricerri and L. Vanhecke, *Generalized Heisenberg groups and Damek-Ricci harmonic spaces*, Lecture Notes in Math. 1598, Springer-Verlag, Berlin, Heidelberg, New York, 1995. MR **97a**:53068
- [4] E. Boeckx and G. Calvaruso, *When is the unit tangent sphere bundle semi-symmetric?*, preprint, 2002.
- [5] E. Boeckx and L. Vanhecke, *Characteristic reflections on unit tangent sphere bundles*, Houston J. Math. **23** (1997), 427–448. MR **2000e**:53052
- [6] E. Boeckx and L. Vanhecke, *Curvature homogeneous unit tangent sphere bundles*, Publ. Math. Debrecen **53** (1998), 389–413. MR **2000d**:53080
- [7] E. Boeckx and L. Vanhecke, *Harmonic and minimal vector fields on tangent and unit tangent bundles*, Differential Geom. Appl. **13** (2000), 77–93. MR **2001f**:53138
- [8] G. de Rham, *Sur la reductibilité d'un espace de Riemann*, Comment. Math. Helv. **26** (1952), 328–344. MR **14**:584a
- [9] O. Kowalski and M. Sekizawa, *On tangent sphere bundles with small or large constant radius*, Ann. Global Anal. Geom. **18** (2000), 207–219. MR **2001i**:53049
- [10] O. Kowalski, M. Sekizawa and Z. Vlášek, *Can tangent sphere bundles over Riemannian manifolds have strictly positive sectional curvature?*, in: Global Differential Geometry: The Mathematical Legacy of Alfred Gray (eds. M. Fernández, J. A. Wolf), Contemp. Math. **288**, Amer. Math. Soc., Providence, RI, 2001, 110–118. MR **2002i**:53047

DEPARTMENT OF MATHEMATICS, KATHOLIEKE UNIVERSITEIT LEUVEN, CELESTIJNENLAAN 200B,
3001 LEUVEN, BELGIUM

E-mail address: eric.boeckx@wis.kuleuven.ac.be

CRITERIA FOR LARGE DEVIATIONS

HENRI COMMAN

ABSTRACT. We give the general variational form of

$$\limsup \left(\int_X e^{h(x)/t_\alpha} \mu_\alpha(dx) \right)^{t_\alpha}$$

for any bounded above Borel measurable function h on a topological space X , where (μ_α) is a net of Borel probability measures on X , and (t_α) a net in $]0, \infty[$ converging to 0. When X is normal, we obtain a criterion in order to have a limit in the above expression for all h continuous bounded, and deduce new criteria of a large deviation principle with not necessarily tight rate function; this allows us to remove the tightness hypothesis in various classical theorems.

1. INTRODUCTION

Let $\mathcal{C}_b(X)$ be the set of real-valued bounded continuous functions on a topological space X , (μ_α) a net of Borel probability measures on X , and (t_α) a net in $]0, \infty[$ converging to 0. For each $[-\infty, +\infty]$ -valued Borel measurable function h on X , we write $\mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$ for $(\int_X e^{h(x)/t_\alpha} \mu_\alpha(dx))^{t_\alpha}$, and define $\Lambda(h) = \log \lim \mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$ provided the limit exists.

The aim of this paper is to clarify the relation between the existence of $\Lambda(h)$ for all $h \in \mathcal{C}_b(X)$, and the one of a large deviation principle for (μ_α) with powers (t_α) . This problem originates from Varadhan's theorem, which states that if X is regular, then such a principle with tight rate function J implies the existence of $\Lambda(h)$ for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying some tail condition (in particular for all $h \in \mathcal{C}_b(X)$), with moreover $\Lambda(h) = \sup_{x \in X} \{h(x) - J(x)\}$. This theorem is a crucial argument in the proofs of all related results; in particular, these also hold under some tightness hypothesis (of the rate function, or of the net $(\mu_\alpha^{t_\alpha})$, i.e., exponential tightness).

We present here a new approach based on a variational representation of the functional $\limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$, and a criterion of existence of $\Lambda(\cdot)$ on $\mathcal{C}_b(X)$ when X is normal. This leads to functional as well as set-theoretic large deviation criteria, which allow us to remove the tightness condition in various basic results of the theory; moreover, the proofs are nonstandard since there are no compactness arguments in the entire paper. Notice that all the results work for general nets of measures and powers, and (except for Section 4 and “(i) \Rightarrow (ii)” in Theorem 3.3 where X is assumed to be normal) for a general topological states space.

Received by the editors January 3, 2002 and, in revised form, November 9, 2002.

2000 *Mathematics Subject Classification*. Primary 60F10.

This work was supported in part by FONDECYT Grant 3010005.

We begin in Section 2 by stating somewhat unusual equivalent definitions of a large deviation principle (Proposition 2.3) which imply the existence of a minimal rate function.

In Section 3, we prove a general variational form of $\limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_Y)$ for any Borel measurable function h on X and any Borel set $Y \subset X$ satisfying the (localized) tail condition of Varadhan's theorem (Theorem 3.1). By the same methods, we obtain a sufficient (and necessary if X is normal) condition for the existence of $\Lambda(h)$ for all $h \in \mathcal{C}_b(X)$, in the spirit of the Portmanteau theorem (Theorem 3.3). A generalized version of Varadhan's theorem without any tightness assumption is a direct consequence (Corollary 3.4).

In Section 4, assuming that X is normal, we look for necessary and sufficient conditions that $\Lambda(\cdot)$ (as a functional on $\mathcal{C}_b(X)$) must satisfy in order to have a large deviation principle. This is obtained in Theorem 4.1, which gives five such conditions; in particular, it states that large deviations occur if and only if $\lim \Lambda(h_i) = \Lambda(h)$ for each increasing net (h_i) in $\mathcal{C}_b(X)$ converging pointwise to $h \in \mathcal{C}_b(X)$. As corollaries, various basic results hold verbatim without tightness assumptions; this is the case for the equivalence between the Laplace and large deviation principles (Corollary 4.2), and for the variational form of a rate function (4.5); a large deviation principle is characterized as a convergence in a narrow space of set-functions much larger than capacities (Remark 4.5); Corollary 4.3 improves Bryc's theorem by weakening the exponential tightness hypothesis; (4.2) gives the infimum of a rate function J on any closed set in terms of $\liminf \mu_\alpha^{t_\alpha}(\cdot)$ (resp. $\limsup \mu_\alpha^{t_\alpha}(\cdot)$), generalizing a well-known expression of J .

2. PRELIMINARIES

Let \mathcal{F} (resp. \mathcal{G}) denotes the set of closed (resp. open) subsets of X . For each $[0, +\infty]$ -valued function f on X we denote by \bar{f} (resp. $\overset{\circ}{f}$) the least upper semi-continuous function on X greater than f (resp. the greatest lower semi-continuous function on X less than f), and define a map $\gamma_f : \mathcal{G} \rightarrow [0, +\infty]$ by $\gamma_f(G) = \sup_{x \in G} f(x)$ for all $G \in \mathcal{G}$.

We collect here some characterizations of positive bounded upper semi-continuous functions that we will use in the sequel.

Lemma 2.1. *There is a bijection between the set of positive bounded upper semi-continuous functions f on X and the set of maps $\gamma : \mathcal{G} \rightarrow [0, +\infty[$ satisfying*

$$(2.1) \quad \gamma\left(\bigcup_{i \in I} G_i\right) = \sup_{i \in I} \gamma(G_i) \quad \text{for all } \{G_i; i \in I\} \subset \mathcal{G},$$

given by the maps

- $\gamma \mapsto f_\gamma(x) = \inf_{G \in \mathcal{G}, x \in G} \gamma(G)$ for all $x \in X$,
- $f \mapsto \gamma_f$.

Moreover, for each positive bounded function f on X , the following properties hold:

- (i) $\bar{f} = f_{\gamma_f}$;
- (ii) \bar{f} is the unique positive upper semi-continuous function h on X satisfying $\gamma_h = \gamma_f$;
- (iii) $\sup_{x \in Y} \bar{f}(x) = \inf_{G \supset Y, G \in \mathcal{G}} \gamma_f(G)$ for all $Y \subset X$;
- (iv) $\bar{f} = \bigvee \{g \in [0, +\infty]^X : \gamma_g = \gamma_f\}$.

Proof. Let f be a positive bounded upper semi-continuous function on X . For each $Y \subset X$ we have $\sup_Y f \leq \sup_{\{f < \sup_Y f + \varepsilon\}} f \leq \sup_Y f + \varepsilon$; since $\{f < \sup_Y f + \varepsilon\}$ is open and contains Y , we obtain

$$(2.2) \quad \sup_Y f = \inf_{G \supset Y, G \in \mathcal{G}} \gamma_f(G),$$

whence $f_{\gamma_f} = f$. We now prove that $\gamma_{f_\gamma} = \gamma$ for all $\gamma : \mathcal{G} \rightarrow [0, +\infty[$ satisfying (2.1); let γ be such a map, $\lambda > 0$, and define $G_\lambda = \bigcup \{G \in \mathcal{G}; \gamma(G) \leq \lambda\}$. For all $0 \leq \nu < \lambda$ and for all $x \in G_\nu$, we have $f_\gamma(x) \leq \gamma(G_\nu) \leq \nu$ so that $\bigcup_{\nu < \lambda} G_\nu \subset \{f_\gamma < \lambda\}$. For all $x \in \{f_\gamma < \lambda\}$ there is $G \ni x$ such that $\gamma(G) < \lambda$, and thus $G \subset G_\nu$ for some $\nu < \lambda$. This shows that $\{f_\gamma < \lambda\} \subset \bigcup_{\nu < \lambda} G_\nu$, and thus $\{f_\gamma < \lambda\} = \bigcup_{\nu < \lambda} G_\nu$, which is an open set, so that f_γ is upper semi-continuous. Let $G \in \mathcal{G}$. Clearly, $\sup_G f_\gamma \leq \gamma(G)$ (with the convention $\sup_\emptyset = 0$). If $\sup_G f_\gamma < \gamma(G)$, then $G \subset \{f_\gamma < \gamma(G) - \varepsilon\}$ for some $\varepsilon > 0$. Since f_γ is upper semi-continuous, $\{f_\gamma < \gamma(G) - \varepsilon\}$ is open with $\{f_\gamma < \gamma(G) - \varepsilon\} \subset G_{\gamma(G) - \varepsilon}$, and since γ is clearly increasing, we obtain $\gamma(G) \leq \gamma(\{f_\gamma < \gamma(G) - \varepsilon\}) \leq \gamma(G_{\gamma(G) - \varepsilon}) \leq \gamma(G) - \varepsilon$, which gives the contradiction. Thus $\gamma(G) = \sup_G f_\gamma$ for all $G \in \mathcal{G}$ and the first assertion is proved.

Let f be a positive bounded function on X . Then γ_f is bounded and satisfies (2.1), so that f_{γ_f} is upper semi-continuous with $f_{\gamma_f} \geq f$. For all positive bounded upper semi-continuous functions $f_1 \geq f$, we have $\gamma_{f_1} \geq \gamma_f$ and so $f_{\gamma_{f_1}} = f_1 \geq f_{\gamma_f} \geq f$, which implies $f_{\gamma_f} = \bar{f}$. This prove (i). If h is a positive upper semi-continuous function on X satisfying $\gamma_h = \gamma_f$, then h is bounded and (ii) follows from (i). Let $Y \subset X$. By (2.2) we have $\sup_Y \bar{f} = \inf_{G \supset Y, G \in \mathcal{G}} \sup_G \bar{f}$, and since $\gamma_{\bar{f}} = \gamma_f$ by (ii), (iii) holds. Let $h = \bigvee \{g \in [0, +\infty[^X; \gamma_g = \gamma_f\}$. It is easy to see that $\gamma_h = \gamma_f$, and since f_{γ_h} is upper semi-continuous with $f_{\gamma_h} \geq h$, we have $\gamma_h = \gamma_{f_{\gamma_h}} = \gamma_f$ whence $h = f_{\gamma_h} = \bar{f}$ by (i). Thus (iv) holds. \square

Definition 2.2. We say that (μ_α) satisfies a *large deviation principle* with powers (t_α) if there is a lower semi-continuous function $J : X \rightarrow [0, \infty]$ such that

$$\limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in F} e^{-J}(x) \leq \sup_{x \in G} e^{-J}(x) \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$. Then, J is called a rate function for $(\mu_\alpha^{t_\alpha})$, which is said to be tight if it has compact level sets.

Notice that in the literature, a large deviation principle is in general defined for nets $(\mu_\epsilon)_{\epsilon > 0}$ or sequences $(\mu_n^{1/n})_{n \in \mathbb{N}^*}$. In the sequel, when we will refer to known results that will be proved again, we will not make this distinction and state them with general nets of measures and powers.

By (2.3), the following proposition shows that the set of rate functions for $(\mu_\alpha^{t_\alpha})$ has a minimal element; it is the only one if X is regular since it is well known ([2], Lemma 4.1.4).

Proposition 2.3. *The following statements are equivalent:*

- (i) (μ_α) satisfies a large deviation principle with powers (t_α) .
- (ii) There is a map $\gamma : \mathcal{G} \rightarrow [0, 1]$ such that
 - (a) $\limsup \mu_\alpha^{t_\alpha}(F) \leq \gamma(G) \leq \liminf \mu_\alpha^{t_\alpha}(G)$ for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.
 - (b) $\gamma(\bigcup_{i \in I} G_i) = \sup_{i \in I} \gamma(G_i)$ for all $\{G_i : i \in I\} \subset \mathcal{G}$.

(iii) There is a function $f : X \rightarrow [0, 1]$ such that

$$\limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in G} f(x) \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.

(iv) There is a function $f : X \rightarrow [0, 1]$ such that

$$\limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in F} f(x) \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.

If (i) holds, then (i) holds with rate function J given by

$$(2.3) \quad e^{-J} = \bigvee \{f \in [0, 1]^X : f \text{ satisfies (iii) (resp. (iv))}\},$$

and

$$(2.4) \quad \gamma_{e^{-J}} = \bigvee \{\gamma \in [0, 1]^{\mathcal{G}} : \gamma \text{ satisfies (ii)}\}.$$

If (ii) holds with γ , then (i) holds with rate function $-\log f_\gamma$ where

$$f_\gamma(x) = \inf_{G \in \mathcal{G}, x \in G} \gamma(G) \quad \text{for all } x \in X.$$

If (iii) (resp. (iv)) holds with f , then (i) holds with rate function $-\log \bar{f}$.

Proof. If (ii) holds with γ , then f_γ is upper semi-continuous and (i) holds with rate function $-\log f_\gamma$ by Lemma 2.1.

If (iii) holds with f , then (ii) holds with γ_f , and so (i) holds with rate function $-\log \bar{f}$, since $f_{\gamma_f} = \bar{f}$ by Lemma 2.1.

If (iv) holds with f , then put $\gamma(G) = \sup_{F \subset G} \sup_F f$ for all $G \in \mathcal{G}$, and notice that γ satisfies (ii). Thus (i) holds with rate function $-\log f_\gamma$. Since $f \leq \bar{f} \leq f_\gamma$, we have

$$\sup_F f \leq \sup_F \bar{f} \leq \sup_G \bar{f} = \sup_G f \leq \sup_G f_\gamma$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$ (the equality follows from Lemma 2.1 (iii)). Thus (i) holds with rate function $-\log \bar{f}$.

If (i) holds, then (ii), (iii) and (iv) hold. The function

$$h = \bigvee \{f \in [0, 1]^X : f \text{ satisfies (iii) (resp. (iv))}\}$$

obviously satisfies (iii) (resp. (iv)); the same for \bar{h} by the preceding discussion, and $h = \bar{h}$ by the definition of h ; put $e^{-J} = \bar{h}$ and obtain (2.3). The map $\gamma_M = \bigvee \{\gamma \in [0, 1]^{\mathcal{G}} : \gamma \text{ satisfies (ii)}\}$ satisfies (ii), and so (i) holds with rate function J given by $e^{-J} = f_{\gamma_M}$. Since $\gamma_{e^{-J}} = \gamma_{f_{\gamma_M}} = \gamma_M$, (2.4) holds. \square

Corollary 2.4. (*Contraction principle*) Let Y be a topological space, and $\pi : X \rightarrow Y$ a continuous function. If (μ_α) satisfies a large deviation principle with powers (t_α) and rate function J^X , then $(\pi[\mu_\alpha])$ satisfies a large deviation principle with powers (t_α) and rate function $J^Y = \bar{l}$ where $l(y) = \inf_{x \in \pi^{-1}(y)} J^X(x)$ for all $y \in Y$.

Proof. Let J^X be a rate function for $(\mu_\alpha^{t_\alpha})$. The relations

$$\begin{aligned} \limsup \pi[\mu_\alpha]^{t_\alpha}(F) &= \limsup \mu_\alpha^{t_\alpha}(\pi^{-1}(F)) \\ &\leq \sup_{\pi^{-1}(F)} e^{-J^X} \leq \sup_{\pi^{-1}(G)} e^{-J^X} \leq \liminf \pi[\mu_\alpha]^{t_\alpha}(G) \end{aligned}$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$ and Proposition 2.3 show that $(\pi[\mu_\alpha])$ satisfies a large deviation principle with powers (t_α) and rate function $J^Y = -\log \bar{f}$ where $f(y) = \sup_{\pi^{-1}(y)} e^{-J^X}$ for all $y \in Y$ (since $\sup_{\pi^{-1}(F)} e^{-J^X} = \sup_F f$). Equivalently, $J^Y = \overset{\circ}{l}$ where $l(y) = \inf_{\pi^{-1}(y)} J^X$ for all $y \in Y$. \square

3. A GENERAL VARIATIONAL FORMULA

Up to now, the only known condition that ensures the existence of $\Lambda(h)$ for all $h \in \mathcal{C}_b(X)$ (and more generally for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying the tail condition

$$(3.1) \quad \lim_{M \rightarrow \infty} \limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_{\{e^h > M\}}) = 0$$

is the existence of a large deviation principle with tight rate function, and $\Lambda(h)$ is expressed in a variational form in terms of this rate function (Varadhan's theorem, [2] Theorem 4.3.1). In this section, we generalize these results in two directions. First, Theorem 3.1 gives the general variational form of $\limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_Y)$ for any Borel set $Y \subset X$ and any Borel measurable function h on X satisfying the localized tail condition (3.2). Next, Theorem 3.3 gives a sufficient condition for the existence of $\Lambda(\cdot)$ on $\mathcal{C}_b(X)$ which is also necessary when X is normal; moreover, the variational form of $\Lambda(\cdot)$ is obtained in terms of any set-function $\gamma \in [0, 1]^{\mathcal{F} \cup \mathcal{G}}$ satisfying the typical in-between inequalities of large deviations (3.23). As a consequence, Varadhan's theorem is generalized in various ways (Corollary 3.2 and Corollary 3.4).

For each map $h : X \rightarrow [-\infty, +\infty]$ we put $F_{\lambda, \varepsilon}^h = \{e^h \in [\lambda - \varepsilon, \lambda + \varepsilon]\}$ and $G_{\lambda, \varepsilon}^h = \{e^h \in]\lambda - \varepsilon, \lambda + \varepsilon[\}$ for all $\lambda \geq 0$ and for all $\varepsilon > 0$.

Theorem 3.1. *For each Borel set $Y \subset X$, and for each $[-\infty, +\infty]$ -valued Borel measurable function h on X satisfying*

$$(3.2) \quad \lim_{M \rightarrow \infty} \limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_{\{e^h > M\} \cap Y}) = 0,$$

we have

$$(3.3) \quad \limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_Y) = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}^h \cap Y)\}$$

$$(3.4) \quad = \sup_{\{x \in Y, \varepsilon > 0 : e^{h(x)} \leq M\}} \{(e^{h(x)} - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(G_{e^{h(x)}, \varepsilon}^h \cap Y)\}$$

for some $M \in [0, +\infty[$. Moreover, $\lim \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_Y)$ exists if

$$\sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \liminf \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}^h \cap Y)\} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}^h \cap Y)\}.$$

Proof. Let Y be any Borel subset of X . Put $g = e^h$, $G_{\lambda, \varepsilon} = G_{\lambda, \varepsilon}^h \cap Y$, $F_{\lambda, \varepsilon} = F_{\lambda, \varepsilon}^h \cap Y$ for all $\lambda \geq 0$ and for all $\varepsilon > 0$. We have

$$\begin{aligned} \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) &\geq \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{F_{\lambda, \varepsilon}}) \\ &\geq (\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}) \end{aligned}$$

for all $\lambda \geq 0$ and for all $\varepsilon > 0$, and so

$$(3.5) \quad \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) \geq \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon})\}.$$

Thus, in order to prove (3.3) and (3.4), we have to prove that for some $M < \infty$,

$$(3.6) \quad \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) \leq \sup_{\{x \in Y, \varepsilon > 0: g(x) \leq M\}} \{(g(x) - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(G_{g(x), \varepsilon})\}.$$

For all $M \geq 0$, for all $N \in \mathbf{N}^*$ and for all $1 \leq j \leq N$, we define

$$F_{M,N,j} = \{g \in [(j-1)M/N, jM/N]\} \cap Y.$$

We have

$$(3.7) \quad \begin{aligned} \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) &\leq \limsup \left(\sum_{j=1}^N \mu_\alpha(g^{1/t_\alpha} 1_{F_{M,N,j}}) + \mu_\alpha(g^{1/t_\alpha} 1_{\{g > M\} \cap Y}) \right)^{t_\alpha} \\ &\leq \max_{1 \leq j \leq N} \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{F_{M,N,j}}) \vee \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{\{g > M\} \cap Y}). \end{aligned}$$

Since

$$\limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{F_{M,N,j}}) \leq \limsup \mu_\alpha^{t_\alpha}(F_{M,N,j}) \|g 1_{F_{M,N,j}}\|,$$

it follows from (3.7) that

$$(3.8) \quad \begin{aligned} &\limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) \\ &\leq \max_{1 \leq j \leq N} \|g 1_{F_{M,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M,N,j}) \vee \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{\{g > M\} \cap Y}). \end{aligned}$$

Let $M \rightarrow \infty$, $N \rightarrow \infty$ in (3.8) and use (3.2) to obtain some $M_0 \in [0, \infty[$ such that

$$(3.9) \quad \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_Y) \leq \liminf_{N \rightarrow \infty} \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j})\}.$$

Thus, to obtain (3.6) it suffices to show

$$(3.10) \quad \begin{aligned} &\liminf_{N \rightarrow \infty} \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j})\} \\ &\leq \sup_{\{x \in Y, \varepsilon > 0: g(x) \leq M_0\}} \{(g(x) - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(G_{g(x), \varepsilon})\}. \end{aligned}$$

If (3.10) does not hold, then there exists $\nu > 0$ such that

$$(3.11) \quad \begin{aligned} &\liminf_{N \rightarrow \infty} \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j})\} \\ &> \sup_{\{x \in Y, \varepsilon > 0: g(x) \leq M_0\}} \{(g(x) + \nu - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(G_{g(x), \varepsilon})\}. \end{aligned}$$

Take $0 < \varepsilon_0 < \nu/2$ in (3.11) and obtain

$$(3.12) \quad \begin{aligned} &\liminf_{N \rightarrow \infty} \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j})\} \\ &> \sup_{\{x \in Y; g(x) \leq M_0\}} \{(g(x) + \varepsilon_0) \limsup \mu_\alpha^{t_\alpha}(G_{g(x), \varepsilon})\}. \end{aligned}$$

But for all $0 \leq \lambda \leq M_0$ and for all $N > M_0/\varepsilon_0$ we have

$$(3.13) \quad (\lambda + \varepsilon_0) \limsup \mu_\alpha^{t_\alpha}(G_{\lambda, \varepsilon_0}) \geq \|g 1_{F_{M_0,N,j_\lambda}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j_\lambda})$$

where j_λ is such that $\lambda \in [(j_\lambda - 1)M_0/N, j_\lambda M_0/N]$ (since $[(j_\lambda - 1)M_0/N, j_\lambda M_0/N] \subset]\lambda - \varepsilon_0, \lambda + \varepsilon_0[$). When λ ranges over $[0, M_0]$, j_λ ranges over $\{j : 1 \leq j \leq N\}$, and (3.13) implies

$$(3.14) \quad \begin{aligned} &\sup_{0 \leq \lambda \leq M_0} \{(\lambda + \varepsilon_0) \limsup \mu_\alpha^{t_\alpha}(G_{\lambda, \varepsilon_0})\} \\ &\geq \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0,N,j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0,N,j})\} \end{aligned}$$

for all $N > M_0/\varepsilon_0$. Notice that for all $N \in \mathbf{N}^*$ and for all $1 \leq j \leq N$, if $F_{M_0, N, j} \neq \emptyset$, then $j = j_{g(x)}$ for some $x \in Y$. Thus it suffices to consider $\lambda \in \{g(x) : x \in Y, g(x) \leq M_0\}$ in the L.H.S. of (3.14), that is,

$$\begin{aligned} & \sup_{\{x \in Y: g(x) \leq M_0\}} \{(g(x) + \varepsilon_0) \limsup \mu_\alpha^{t_\alpha}(G_{g(x), \varepsilon_0})\} \\ & \geq \max_{1 \leq j \leq N} \{\|g1_{F_{M_0, N, j}}\| \limsup \mu_\alpha^{t_\alpha}(F_{M_0, N, j})\} \end{aligned}$$

for all $N > M_0/\varepsilon_0$, which contradicts (3.12); it follows that (3.10), (3.6), and finally (3.3) and (3.4) hold. In the same way that we obtained (3.5), we have

$$(3.15) \quad \liminf \mu_\alpha^{t_\alpha}(g^{1/t_\alpha}1_Y) \geq \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \liminf \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon})\},$$

and the last assertion follows from (3.4). \square

A localized version of Varadhan's theorem states that if X is regular and if (μ_α) satisfies a large deviation principle with powers (t_α) and tight rate function J , then (3.18) and (3.19) hold with $l = J$ ([2], Exercise 4.3.11). The following corollary removes all the hypotheses on l and X .

Corollary 3.2. *Let l be a $[0, +\infty]$ -valued function on X satisfying*

$$(3.16) \quad \forall F \in \mathcal{F}, \quad \limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in F} e^{-l(x)}$$

$$(3.17) \quad (\text{resp. } \forall G \in \mathcal{G}, \quad \liminf \mu_\alpha^{t_\alpha}(G) \geq \sup_{x \in G} e^{-l(x)}).$$

Then, for each $[-\infty, +\infty]$ -valued continuous function h on X satisfying (3.1), we have

$$(3.18) \quad \forall F \in \mathcal{F}, \quad \limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha}1_F) \leq \sup_{x \in F, h(x) < \infty} e^{h(x)} e^{-l(x)}$$

$$(3.19) \quad (\text{resp. } \forall G \in \mathcal{G}, \quad \liminf \mu_\alpha^{t_\alpha}(e^{h/t_\alpha}1_G) \geq \sup_{x \in G, h(x) < \infty} e^{h(x)} e^{-l(x)}).$$

Proof. Suppose that (3.16) holds and (3.18) does not hold for some $[-\infty, +\infty]$ -valued continuous function h on X satisfying (3.1). Since

$$\limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha}1_F) = \sup_{\{x \in F, \varepsilon > 0: h(x) < \infty\}} \{(e^{h(x)} - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x)}, \varepsilon}^h \cap F)\}$$

for all $F \in \mathcal{F}$ by Theorem 3.1, there exists $F_0 \in \mathcal{F}$, $x_0 \in F_0$ with $h(x_0) < \infty$, and $\varepsilon_0 > 0$ such that

$$(e^{h(x_0)} - \varepsilon_0) \limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x_0)}, \varepsilon_0}^h \cap F) > \sup_{x \in F, h(x) < \infty} e^{h(x)} e^{-l(x)}.$$

By (3.16) we have

$$(e^{h(x_0)} - \varepsilon_0) \sup_{x \in F_{e^{h(x_0)}, \varepsilon_0}^h \cap F} e^{-l(x)} > \sup_{x \in F, h(x) < \infty} e^{h(x)} e^{-l(x)},$$

and so there exists $x_1 \in F_{e^{h(x_0)}, \varepsilon_0}^h \cap F$ such that

$$(e^{h(x_0)} - \varepsilon_0) e^{-l(x_1)} > \sup_{x \in F, h(x) < \infty} e^{h(x)} e^{-l(x)}.$$

Since $e^{h(x_1)} \geq e^{h(x_0)} - \varepsilon_0$ we obtain

$$e^{h(x_1)} e^{-l(x_1)} > \sup_{x \in F, h(x) < \infty} e^{h(x)} e^{-l(x)}$$

with $x_1 \in F$ and $h(x_1) < \infty$, whence the contradiction. Suppose now that (3.17) holds and (3.19) does not hold for some $[-\infty, +\infty]$ -valued continuous function h on X satisfying (3.1). By (3.15), there exists $G_0 \in \mathcal{G}$ such that

$$\begin{aligned} & \sup_{x \in G_0, h(x) < \infty} e^{h(x)} e^{-l(x)} > \liminf \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_{G_0}) \\ & \geq \sup_{\{x \in G_0, \varepsilon > 0: h(x) < \infty\}} \{(e^{h(x)} - \varepsilon) \liminf \mu_\alpha^{t_\alpha}(G_{e^{h(x)}, \varepsilon}^h \cap G_0)\}. \end{aligned}$$

Thus, there exists $x_0 \in G_0$ with $h(x_0) < \infty$, and $\nu > 0$ such that

$$e^{h(x_0)} e^{-l(x_0)} > \nu + \sup_{\{x \in G_0, \varepsilon > 0: h(x) < \infty\}} \{(e^{h(x)} - \varepsilon) \liminf \mu_\alpha^{t_\alpha}(G_{e^{h(x)}, \varepsilon}^h \cap G_0)\}$$

and

$$(3.20) \quad e^{h(x_0)} e^{-l(x_0)} > \sup_{\{x \in G_0, \varepsilon > 0: h(x) < \infty\}} \{(e^{h(x)} - \varepsilon + \nu) \liminf \mu_\alpha^{t_\alpha}(G_{e^{h(x)}, \varepsilon}^h \cap G_0)\}.$$

By taking $x = x_0$ and $\varepsilon_0 < \nu$ in the R. H. S. of (3.20) we have

$$e^{h(x_0)} e^{-l(x_0)} > e^{h(x_0)} \liminf \mu_\alpha^{t_\alpha}(G_{e^{h(x_0)}, \varepsilon_0}^h \cap G_0),$$

and by (3.17)

$$e^{h(x_0)} e^{-l(x_0)} > e^{h(x_0)} \sup_{x \in G_{e^{h(x_0)}, \varepsilon_0}^h \cap G} e^{-l(x)},$$

which gives the contradiction. \square

A direct consequence of Corollary 3.2 is that Varadhan's theorem can be stated verbatim for a general state space and with any function (in place of a tight rate function) $l : X \rightarrow [0, +\infty]$ satisfying the large deviations lower and upper bounds:

$$(3.21) \quad \limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in F} e^{-l(x)} \leq \sup_{x \in G} e^{-l(x)} \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$; that is to say, $\Lambda(h)$ exists and

$$\Lambda(h) = \sup_{x \in X, h(x) < \infty} \{h(x) - l(x)\}$$

for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying (3.1). We will see with Corollary 3.4 that it is possible to go further in the generalization of Varadhan's theorem obtaining the same conclusions with hypothesis weaker than (3.21).

Recall that X is normal if and only if the following interpolation property holds: if f and g are real-valued respectively upper and lower semi-continuous functions on X such that $f \leq g$, then there is a continuous function h on X satisfying $f \leq h \leq g$.

Theorem 3.3. *Consider the following statements:*

- (i) $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$;
- (ii) $\limsup \mu_\alpha^{t_\alpha}(F) \leq \liminf \mu_\alpha^{t_\alpha}(G)$ for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.

If X is normal, then (i) \Rightarrow (ii). If (ii) holds, then (i) holds and moreover for each $[-\infty, +\infty]$ -valued continuous function h on X satisfying (3.1) we have for some $M \in [0, +\infty[$,

$$(3.22) \quad e^{\Lambda(h)} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \gamma(F_{\lambda, \varepsilon}^h)\} = \sup_{\{x \in X, \varepsilon > 0: e^{h(x)} \leq M\}} \{(e^{h(x)} - \varepsilon) \gamma(G_{e^{h(x)}, \varepsilon}^h)\}$$

for all maps $\gamma : \mathcal{F} \cup \mathcal{G} \rightarrow [0, 1]$ satisfying

$$(3.23) \quad \limsup \mu_\alpha^{t_\alpha}(F) \leq \gamma(F) \leq \gamma(G) \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.

Proof. Suppose that (i) holds and X is normal. For each $F \in \mathcal{F}$ and $G \in \mathcal{G}$ with $F \subset G$, there exists $h \in \mathcal{C}_b(X)$ such that $1_F \leq h \leq 1_G$. Since $1_F \leq e^{nh-n} \leq e^{-n} 1_{X \setminus G} \vee 1_G$ for all $n \in \mathbb{N}$, we obtain

$$\begin{aligned} \limsup \mu_\alpha^{t_\alpha}(F) &\leq \inf_n e^{\Lambda(nh-n)} \\ &\leq \inf_n \liminf \{e^{-n} + \mu_\alpha^{t_\alpha}(G)\} \leq \liminf \mu_\alpha^{t_\alpha}(G) \end{aligned}$$

and (ii) holds.

Suppose that (ii) holds. Let h be a $[-\infty, +\infty]$ -valued continuous function on X satisfying (3.1), and $\gamma : \mathcal{F} \cup \mathcal{G} \rightarrow [0, 1]$ satisfying (3.23). Put $g = e^h$ and let us use the same notation as in the proof of Theorem 3.1 (with $Y = X$). For all $\lambda \geq 0$, for all $\varepsilon > 0$ and for all $\delta > 0$ with $\delta > \varepsilon$, we have by (3.23),

$$\begin{aligned} \liminf \mu_\alpha^{t_\alpha}(g^{1/t_\alpha}) &\geq \liminf \mu_\alpha^{t_\alpha}(g^{1/t_\alpha} 1_{G_{\lambda,\delta}}) \\ &\geq (\lambda - \delta) \gamma(G_{\lambda,\delta}) \geq (\lambda - \delta) \gamma(F_{\lambda,\varepsilon}). \end{aligned}$$

Thus

$$\begin{aligned} \liminf \mu_\alpha^{t_\alpha}(g^{1/t_\alpha}) &\geq \lim_{\delta \rightarrow \varepsilon} (\lambda - \delta) \gamma(F_{\lambda,\varepsilon}) \\ &\geq (\lambda - \varepsilon) \gamma(F_{\lambda,\varepsilon}) \geq (\lambda - \varepsilon) \gamma(G_{\lambda,\varepsilon}) \end{aligned}$$

and

$$\begin{aligned} \liminf \mu_\alpha^{t_\alpha}(g^{1/t_\alpha}) &\geq \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \gamma(F_{\lambda,\varepsilon})\} \\ &\geq \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \gamma(G_{\lambda,\varepsilon})\}. \end{aligned}$$

In order to prove (3.22), we have to prove that for some $M < \infty$,

$$(3.24) \quad \limsup \mu_\alpha^{t_\alpha}(g^{1/t_\alpha}) \leq \sup_{\{x \in X, \varepsilon > 0: e^{h(x)} \leq M\}} \{(e^{h(x)} - \varepsilon) \gamma(G_{e^{h(x)}, \varepsilon})\}.$$

By using (3.23), and in the same way that we have obtained (3.10) in the proof of Theorem 3.1, we find some $M_0 \in [0, \infty[$ such that to prove (3.24) it suffices to prove

$$\liminf_{N \rightarrow \infty} \max_{1 \leq j \leq N} \{\|g 1_{F_{M_0, N, j}}\| \gamma(F_{M_0, N, j})\} \leq \sup_{\{x \in X, \varepsilon > 0: e^{h(x)} \leq M_0\}} \{(e^{h(x)} - \varepsilon) \gamma(G_{e^{h(x)}, \varepsilon})\},$$

which is achieved exactly as in Theorem 3.1. \square

The following corollary gives sufficient conditions much weaker than large deviations with tight rate function in order to have the conclusions of Varadhan's theorem; in fact, we will see in the next section (Corollary 4.2) that when X is normal, the condition (3.25) is also necessary.

Corollary 3.4. *Let l be a $[0, +\infty]$ -valued function on X satisfying*

$$(3.25) \quad \limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in G} e^{-l(x)} \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

$$(3.26) \quad (\text{resp. } \limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in F} e^{-l(x)} \leq \liminf \mu_\alpha^{t_\alpha}(G))$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$. Then, $\Lambda(h)$ exists and

$$(3.27) \quad \Lambda(h) = \sup_{x \in X, h(x) < \infty} \{h(x) - l(x)\}$$

for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying (3.1).

Proof. Let h be a $[-\infty, +\infty]$ -valued continuous function on X satisfying (3.1). If (3.25) holds, then by Theorem 3.3 (with $\gamma(G) = \sup_{x \in G} e^{-l(x)}$ for all $G \in \mathcal{G}$), $\Lambda(h)$ exists and

$$e^{\Lambda(h)} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \sup_{x \in G_{\lambda, \varepsilon}^h} e^{-l(x)}\}.$$

Since for all $\lambda \geq 0$, $\varepsilon > 0$ and $x \in G_{\lambda, \varepsilon}^h$,

$$(\lambda - \varepsilon)e^{-l(x)} \leq e^{h(x)}e^{-l(x)},$$

we obtain

$$\begin{aligned} (\lambda - \varepsilon)e^{-l(x)} &\leq \sup_{x \in X, h(x) < \infty} e^{h(x)}e^{-l(x)}, \\ (\lambda - \varepsilon) \sup_{x \in G_{\lambda, \varepsilon}^h} e^{-l(x)} &\leq \sup_{x \in X, h(x) < \infty} e^{h(x)}e^{-l(x)}, \end{aligned}$$

and thus

$$e^{\Lambda(h)} \leq \sup_{x \in X, h(x) < \infty} e^{h(x)}e^{-l(x)}.$$

For all $x \in X$ with $h(x) < \infty$, and for all $\varepsilon > 0$ we have

$$(e^{h(x)} - \varepsilon)e^{-l(x)} \leq (e^{h(x)} - \varepsilon) \sup_{y \in G_{e^{h(x)}, \varepsilon}^h} e^{-l(y)},$$

which implies

$$\begin{aligned} (e^{h(x)} - \varepsilon)e^{-l(x)} &\leq e^{\Lambda(h)}, \\ e^{h(x)}e^{-l(x)} &\leq e^{\Lambda(h)}, \end{aligned}$$

and finally

$$\sup_{x \in X, h(x) < \infty} e^{h(x)}e^{-l(x)} \leq e^{\Lambda(h)}.$$

Thus $e^{\Lambda(h)} = \sup_{x \in X, h(x) < \infty} e^{h(x)}e^{-l(x)}$, which is equivalent to (3.27). If (3.26) holds, we conclude by applying Theorem 3.3 (with $\gamma(F) = \sup_{x \in F} e^{-l(x)}$ for all $F \in \mathcal{F}$), and replacing $G_{\lambda, \varepsilon}^h$ by $F_{\lambda, \varepsilon}^h$, and $G_{e^{h(x)}, \varepsilon}^h$ by $F_{e^{h(x)}, \varepsilon}^h$ in the above proof. \square

Remark 3.5. Let Γ be the set of maps $\gamma : \mathcal{F} \cup \mathcal{G} \rightarrow [0, 1]$ such that $\gamma(F) \leq \gamma(G)$ for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$. Define the narrow topology on Γ as the coarsest topology for which the mappings $\gamma \mapsto \gamma(Y)$ are upper semi-continuous for all $Y \in \mathcal{F}$ and lower semi-continuous for all $Y \in \mathcal{G}$. The net $(\mu_\alpha^{t_\alpha}(\cdot^{1/t_\alpha}))$ can be seen as a net in Γ provided with the narrow topology, as well as a net in $[0, \infty[^{\{e^h : h \in \mathcal{C}_b(X)\}}$ provided with the product topology. Then, the implication (ii) \Rightarrow (i) in Theorem 3.3 means that if $(\mu_\alpha^{t_\alpha}(\cdot^{1/t_\alpha}))$ has a limit in Γ , then $(\mu_\alpha^{t_\alpha}(\cdot^{1/t_\alpha}))$ has a limit in $[0, \infty[^{\{e^h : h \in \mathcal{C}_b(X)\}}$; moreover, the converse holds if X is normal. Of course, the limit in Γ when it exists is not unique: for each $F \in \mathcal{F}$ and $G \in \mathcal{G}$, γ defined by $\gamma(F) = \limsup \mu_\alpha^{t_\alpha}(F)$ and $\gamma(G) = \liminf \mu_\alpha^{t_\alpha}(G)$ is an example, and γ' defined by $\gamma'(G) = \gamma(G)$ and $\gamma'(F) = \inf_{G \supset F, G \in \mathcal{G}} \gamma(G)$ is another one.

4. CRITERIA OF A LARGE DEVIATION PRINCIPLE

In this section, we investigate what has to be added to the existence of $\Lambda(h)$ for all $h \in \mathcal{C}_b(X)$ (in other words, of the limit $\Lambda(\cdot)$ of $(\log \mu_\alpha^{t_\alpha}(e^{\cdot/t_\alpha}))$ in $]-\infty, +\infty[^{\mathcal{C}_b(X)}$) in order to have large deviations. Of course, some hypotheses on X are required to have sufficiently continuous functions; so we suppose here that X is normal. In this case, by Theorem 3.3 (and Remark 3.5) the existence of $\Lambda(\cdot)$ on $\mathcal{C}_b(X)$ is equivalent to the existence of a narrow set-theoretic limit $\gamma \in \Gamma$ of $(\mu_\alpha^{t_\alpha})$, which is also equivalent to the existence of $\Lambda(h)$ for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying the tail condition (3.1); moreover, for each such function h , the variational form of $\Lambda(h)$ is given in terms of γ . In particular, γ can vary and it is essentially this flexibility which allows us to obtain in Theorem 4.1 necessary and sufficient conditions, each of them corresponding to some type of information: a property of $\Lambda(\cdot)$ as a functional in (ii), a special variational form of $\Lambda(\cdot)$ in (iii), a property of γ in (iv), and a property of the net $(\mu_\alpha^{t_\alpha})$ in (v) and (vi). It is worth noticing that in both formulations (functional (ii) or set-theoretic (iv)), the condition on the limit is the same: a continuity property on increasing nets. As corollaries, several basic results of the theory are strengthened by removing the tightness or compactness hypothesis.

Theorem 4.1. *If X is normal, then the following statements are equivalent:*

- (i) (μ_α) satisfies a large deviation principle with powers (t_α) .
- (ii) $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$, and $\Lambda(h_i)$ converges to $\Lambda(h)$ for each increasing net (h_i) in $\mathcal{C}_b(X)$ converging pointwise to $h \in \mathcal{C}_b(X)$.
- (iii) $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$, and $\Lambda(h) = \sup_{x \in X} \{h(x) - l(x)\}$ for some function $l : X \rightarrow [0, +\infty]$ and for all $h \in \mathcal{C}_b(X)$.
- (iv) There is a map $\gamma : \mathcal{G} \rightarrow [0, 1]$ such that
 - (a) $\limsup \mu_\alpha^{t_\alpha}(F) \leq \gamma(G) \leq \liminf \mu_\alpha^{t_\alpha}(G)$ for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.
 - (b) $\gamma(\bigcup_i G_i) = \lim \gamma(G_i)$ for each increasing net (G_i) in \mathcal{G} .
- (v) $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$, and for all $F \in \mathcal{F}$, for all open covers $\{G_i : i \in I\}$ of F and for all $\varepsilon > 0$, there exists a finite subset $\{G_{i_1}, \dots, G_{i_N}\} \subset \{G_i : i \in I\}$ such that

$$(4.1) \quad \liminf \mu_\alpha^{t_\alpha}(\overset{\circ}{F}) - \limsup \mu_\alpha^{t_\alpha}(\bigcup_{1 \leq j \leq N} \overline{G_{i_j}}) < \varepsilon.$$

- (vi) There is a function $l : X \rightarrow [0, +\infty]$ such that

$$(4.2) \quad \inf_{x \in F} l(x) = \sup_{G \in \mathcal{G}, G \supset F} \{-\liminf t_\alpha \log \mu_\alpha(G)\} = \sup_{G \in \mathcal{G}, G \supset F} \{-\limsup t_\alpha \log \mu_\alpha(G)\}$$

for all $F \in \mathcal{F}$.

If (i) holds with rate function J , then the following properties hold:

$$(4.3) \quad \inf_{x \in F} J(x) = \sup_{h \in \mathcal{C}_b(X), h|_F = 0} \{-\Lambda(h)\} \quad \text{for all } F \in \mathcal{F};$$

$$(4.4) \quad \inf_{x \in G} J(x) = \sup_{h \in \mathcal{C}_{b_a}(X), e^h \leq 1_G} \{-\Lambda(h)\} \quad \text{for all } G \in \mathcal{G},$$

where $C_{ba}(X)$ is the set of $[-\infty, +\infty[$ -valued bounded above continuous functions on X ; in particular,

$$(4.5) \quad J(x) = \sup_{h \in \mathcal{C}_b(X)} \{h(x) - \Lambda(h)\} \quad \text{for all } x \in X;$$

$$(4.6) \quad J = \overset{\circ}{l} \quad \text{for all } l : X \rightarrow [0, +\infty] \text{ satisfying (iii);}$$

$$(4.7) \quad e^{-J}(x) = \inf_{G \in \mathcal{G}, x \in G} \gamma(G)$$

for all $x \in X$, and for all $\gamma : \mathcal{G} \rightarrow [0, 1]$ satisfying (iv);

$$(4.8) \quad J = \overset{\circ}{l} \quad \text{for all } l : X \rightarrow [0, +\infty] \text{ satisfying (v).}$$

If moreover X is second countable, then we can replace “net” by “sequence” in (ii) (resp. (iv)), and “open covers” by “countable open covers” in (v).

Proof. (i) \Rightarrow (iv) and (iii) \Rightarrow (ii) are clear; (i) \Rightarrow (iii) by Corollary 3.4 and so (i) \Rightarrow (ii). If (i) holds with rate function J , then for each $F \in \mathcal{F}$, each open cover $\{G_i : i \in I\}$ of F and each $\varepsilon > 0$,

$$\limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_F e^{-J} \leq \sup_{\bigcup_{i \in I} G_i} e^{-J} = \sup_{i \in I} \sup_{G_i} e^{-J} < \sup_{i \in I} \liminf \mu_\alpha^{t_\alpha}(G_i) + \varepsilon,$$

which implies (v).

Suppose that (ii) holds. We will prove that (i) holds. Let $C_{ba}(X)$ be the set of $[-\infty, +\infty[$ -valued bounded above continuous functions on X . By Theorem 3.3, $\Lambda(h)$ exists in $[-\infty, +\infty[$ for all $h \in \mathcal{C}_{ba}(X)$, and notice that

$$\Lambda(h \vee k) = \Lambda(h) \vee \Lambda(k)$$

for all $k \in \mathcal{C}_{ba}(X)$; in particular,

$$\Lambda(h \vee s) = \Lambda(h) \vee s$$

for all $s \in [-\infty, +\infty[$. Let (h_i) be an increasing net in $\mathcal{C}_{ba}(X)$ converging to $h \in \mathcal{C}_{ba}(X)$ with $\Lambda(h) > -\infty$. For each real $s < \Lambda(h)$ we have $\lim \Lambda(h_i \vee s) = \Lambda(h \vee s) = \Lambda(h)$, and so eventually $\Lambda(h_i) > s$, which shows that $\lim \Lambda(h_i) = \Lambda(h)$. Therefore, we can replace $\mathcal{C}_b(X)$ by $\mathcal{C}_{ba}(X)$ in (ii). Let $F \in \mathcal{F}$ and $h \in \mathcal{C}_{ba}(X)$ with $h|_F = 0$. If $\Lambda(h) > -\infty$, then $\Lambda(h \vee s) = \Lambda(h)$ with $(h \vee s)|_F = 0$ for all $s < \Lambda(h) \wedge 0$; if $\Lambda(h) = -\infty$, then the sequence $(\Lambda(h \vee -n))_{n \in \mathbf{N}}$ converges to $-\infty$ with $(h \vee -n)|_F = 0$. Thus,

$$\inf_{h \in \mathcal{C}_{ba}(X), h|_F=0} e^{\Lambda(h)} = \inf_{h \in \mathcal{C}_b(X), h|_F=0} e^{\Lambda(h)},$$

and by the interpolation property we have

$$(4.9) \quad \begin{aligned} \limsup \mu_\alpha^{t_\alpha}(F) &\leq \inf_{h \in \mathcal{C}_{ba}(X), h|_F=0} e^{\Lambda(h)} = \inf_{h \in \mathcal{C}_b(X), h|_F=0} e^{\Lambda(h)} \\ &\leq \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)} \leq \liminf \mu_\alpha^{t_\alpha}(G) \end{aligned}$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$. Define

$$f(x) = \inf_{h \in \mathcal{C}_b(X), h(x)=0} e^{\Lambda(h)} \quad (= \inf_{h \in \mathcal{C}_{ba}(X), h(x)=0} e^{\Lambda(h)})$$

for all $x \in X$. By (4.9), in order to prove (i) it suffices to show that f is upper semi-continuous and satisfies

$$(4.10) \quad \sup_{x \in F} f(x) = \inf_{h \in \mathcal{C}_b(X), h|_F = 0} e^{\Lambda(h)}$$

for all $F \in \mathcal{F}$ and

$$(4.11) \quad \sup_{x \in G} f(x) = \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

for all $G \in \mathcal{G}$. We first show (4.10). Clearly

$$\sup_{x \in F} f(x) \leq \inf_{h \in \mathcal{C}_b(X), h|_F = 0} e^{\Lambda(h)}$$

for all $F \in \mathcal{F}$. Suppose that

$$\sup_{x \in F} f(x) < e^s < \inf_{h \in \mathcal{C}_b(X), h|_F = 0} e^{\Lambda(h)}$$

for some $F \in \mathcal{F}$ and some real s . Then, for all $x \in F$ there exists $h_x \in \mathcal{C}_b(X)$ which can be chosen negative such that $h_x(x) = 0$ and

$$(4.12) \quad \Lambda(h_x) < s < \inf_{h \in \mathcal{C}_b(X), h|_F = 0} \Lambda(h).$$

But $1_F \leq e^{\bigvee_{x \in F} h_x}$ with $\bigvee_{x \in F} h_x$ bounded lower semi-continuous, and so there exists $h \in \mathcal{C}_b(X)$ such that $1_F \leq e^h \leq e^{\bigvee_{x \in F} h_x}$ (in particular $h|_F = 0$). Let I be the set of finite subsets of F ordered by inclusion, and $h_i = h \wedge \bigvee_{x \in i} h_x$ for all $i \in I$, so that $(h_i)_{i \in I}$ is an increasing net in $\mathcal{C}_b(X)$ converging to h . Since $\Lambda(h_i) \leq \Lambda(\bigvee_{x \in i} h_x) = \sup_{x \in i} \Lambda(h_x) < s$ for all $i \in I$, we obtain $\lim \Lambda(h_i) = \Lambda(h) \leq s$, which contradicts (4.12). Thus (4.10) holds. We now prove (4.11). By the interpolation property (between $1_{\{x\}}$ and 1_G) we have clearly

$$\sup_{x \in G} f(x) \leq \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

for all $G \in \mathcal{G}$. Suppose

$$\sup_{x \in G} f(x) < \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

for some $G \in \mathcal{G}$. Then, for all $x \in G$ there exists $h_x \in \mathcal{C}_b(X)$ with $h_x(x) = 0$ such that

$$(4.13) \quad \sup_{x \in G} \Lambda(h_x) < s < \Lambda(h_G)$$

for some $h_G \in \mathcal{C}_{ba}(X)$ with $e^{h_G} \leq 1_G$, and some real s . Let I be the set of finite subsets of G ordered by inclusion, and $h_i = h_G \wedge \bigvee_{x \in i} h_x$ for all $i \in I$, so that $(h_i)_{i \in I}$ is an increasing net in $\mathcal{C}_{ba}(X)$ converging to h_G . Then

$$\Lambda(h_G) = \lim \Lambda(h_i) \leq \lim \Lambda\left(\bigvee_{x \in i} h_x\right) = \lim(\sup_{x \in i} \Lambda(h_x)) \leq s,$$

which contradicts (4.13). Thus (4.11) holds. It remains to show that f is upper semi-continuous. By (4.9), (4.10), (4.11), and since

$$f(x) = \inf_{h \in \mathcal{C}_{ba}(X), h(x) = 0} e^{\Lambda(h)} \leq \inf_{G \supset \{x\}} \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

for all $x \in X$, by Lemma 2.1 it suffices to prove that

$$(4.14) \quad f(x) = \inf_{h \in C_{ba}(X), h(x)=0} e^{\Lambda(h)} \geq \inf_{G \supset \{x\}} \sup_{h \in C_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

for all $x \in X$. Suppose that (4.14) does not hold for some $x \in X$. Then, there exists $h_x \in C_{ba}(X)$ with $h_x(x) = 0$, and $\nu > 0$ such that

$$(4.15) \quad e^{\Lambda(h_x)} + \nu < \inf_{G \supset \{x\}} \sup_{h \in C_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}.$$

By (4.9) and Theorem 3.3 (with $\gamma(G) = \sup_{h \in C_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$), we have

$$(4.16) \quad e^{\Lambda(h_x)} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \sup_{h \in C_{ba}(X), e^h \leq 1_{G_{\lambda, \varepsilon}^{h_x}}} e^{\Lambda(h)}\}.$$

Take $\lambda = 1$ and $0 < \varepsilon < \nu$ in (4.16), and obtain by (4.15),

$$\sup_{h \in C_{ba}(X), e^h \leq 1_{G_{1, \varepsilon}^{h_x}}} e^{\Lambda(h)} < \inf_{G \supset \{x\}} \sup_{h \in C_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)}$$

with $x \in G_{1, \varepsilon}^{h_x}$, which gives the contradiction. Thus (4.14) holds and f is upper semi-continuous.

We have proved (i) \Leftrightarrow (ii), and that when (i) holds with rate function J , then (4.3) and (4.4) hold (by the uniqueness of a rate function on regular spaces); since

$$\Lambda(h - h(x)) = \Lambda(h) - h(x),$$

(4.5) follows from (4.3).

Suppose that (iii) holds with $l : X \rightarrow [0, +\infty]$. Then, obviously (ii) and so (i) hold; let J be the associated rate function. By Corollary 3.4,

$$(4.17) \quad e^{\Lambda(h)} = \sup_X e^h e^{-J}$$

for all $h \in C_{ba}(X)$, and so

$$(4.18) \quad \sup_X e^h e^{-J} = \sup_X e^h e^{-l}$$

for all $h \in C_b(X)$. Clearly, for each $h \in C_{ba}(X)$ there exists a real s such that

$$\sup_X e^h e^{-J} = \sup_X e^{h \vee s} e^{-J}$$

and

$$\sup_X e^h e^{-l} = \sup_X e^{h \vee s} e^{-l},$$

and so by (4.18),

$$(4.19) \quad \sup_X e^h e^{-J} = \sup_X e^h e^{-l}.$$

For any $G \in \mathcal{G}$ choose an increasing net (h_i) in $C_{ba}(X)$ such that $\sup_i e^{h_i} = 1_G$, and obtain by (4.19),

$$\sup_G e^{-J} = \sup_G e^{-l}.$$

Since $\sup_G e^{-l} = \sup_G e^{-\overset{\circ}{l}}$ by Lemma 2.1, we have $\sup_G e^{-J} = \sup_G e^{-\overset{\circ}{l}}$ for all $G \in \mathcal{G}$. Since e^{-J} and $e^{-\overset{\circ}{l}}$ are upper semi-continuous, we have $J = \overset{\circ}{l}$ by Lemma 2.1. Thus, if (i) holds with rate function J , then $J = \overset{\circ}{l}$ for all $l : X \rightarrow [0, +\infty]$ satisfying (iii) and (4.6) holds.

Suppose that (iv) holds with $\gamma : \mathcal{G} \rightarrow [0, 1]$. Define $\gamma(F) = \inf_{G \supset F} \gamma(G)$ for all $F \in \mathcal{F}$, and notice that γ is increasing on \mathcal{F} , satisfies $\gamma(F) \leq \gamma(G)$ for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$, and by (a),

$$(4.20) \quad \gamma\left(\bigcup_{1 \leq j \leq N} G_j\right) \leq \sup_{1 \leq j \leq N} \gamma(\overline{G_j})$$

for each finite family $\{G_j\}_{1 \leq j \leq N} \subset \mathcal{G}$. By Theorem (3.3), $\Lambda(h)$ exists for all $h \in \mathcal{C}_{ba}(X)$ and

$$(4.21) \quad e^{\Lambda(h)} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon)\gamma(G_{\lambda, \varepsilon}^h)\} = \sup_{x \in X, \varepsilon > 0} \{(e^{h(x)} - \varepsilon)\gamma(G_{e^{h(x)}, \varepsilon}^h)\}$$

$$(4.22) \quad = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon)\gamma(F_{\lambda, \varepsilon}^h)\}.$$

We will show that (ii) holds. Let $(h_i)_{i \in I}$ be an increasing net in $\mathcal{C}_b(X)$ converging to $h \in \mathcal{C}_b(X)$, and suppose that $\Lambda(h) > \sup_{i \in I} \Lambda(h_i)$. By (4.21) and (4.22), there exists $\lambda_0 \geq 0$ and $\varepsilon_0 > 0$ such that

$$\begin{aligned} (\lambda_0 - \varepsilon_0)\gamma(G_{\lambda_0, \varepsilon_0}^h) &> \sup_{i \in I} \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon)\gamma(F_{\lambda, \varepsilon}^{h_i})\} \\ &\geq \sup_{i \in I} \{(\lambda_0 - \varepsilon_0)\gamma(F_{\lambda_0, \varepsilon_0}^{h_i})\}, \end{aligned}$$

and thus

$$\gamma(G_{\lambda_0, \varepsilon_0}^h) > \sup_{i \in I} \gamma(F_{\lambda_0, \varepsilon_0}^{h_i}) \geq \sup_{i \in I} \gamma(\overline{G_{\lambda_0, \varepsilon_0}^{h_i}}).$$

Let \wp be the set of finite subsets of I ordered by inclusion, and obtain by (4.20),

$$(4.23) \quad \forall \beta \in \wp, \quad \gamma(G_{\lambda_0, \varepsilon_0}^h) > \sup_{\beta \in \wp} \gamma\left(\bigcup_{i \in \beta} G_{\lambda_0, \varepsilon_0}^{h_i}\right).$$

But $G_{\lambda_0, \varepsilon_0}^h \subset \sup_{\beta \in \wp} \bigcup_{i \in \beta} G_{\lambda_0, \varepsilon_0}^{h_i}$, and the condition (b) contradicts (4.23). It follows that $\Lambda(h) = \sup_{i \in I} \Lambda(h_i)$, that is, (ii) and so (i) hold; let J be the associated rate function. We now prove that (4.7) holds. Let $G \in \mathcal{G}$ and $h \in \mathcal{C}_{ba}(X)$ be such that $e^h \leq 1_G$. For all $x \in X$ and $\varepsilon > 0$ with $e^{h(x)} > \varepsilon$, we have

$$F_{e^{h(x)}, \varepsilon}^h \subset G$$

and

$$(e^{h(x)} - \varepsilon)\gamma(F_{e^{h(x)}, \varepsilon}^h) \leq (e^{h(x)} - \varepsilon)\gamma(G).$$

Thus,

$$(4.24) \quad \sup_{\{x \in X; e^{h(x)} > \varepsilon\}, \varepsilon > 0} \{(e^{h(x)} - \varepsilon)\gamma(F_{e^{h(x)}, \varepsilon}^h)\} \leq \sup_{\{x \in X; e^{h(x)} > \varepsilon\}, \varepsilon > 0} \{(e^{h(x)} - \varepsilon)\gamma(G)\} \leq \gamma(G),$$

and since if $e^{\Lambda(h)} > 0$,

$$(4.25) \quad e^{\Lambda(h)} = \sup_{\{x \in X; e^{h(x)} > \varepsilon\}, \varepsilon > 0} \{(e^{h(x)} - \varepsilon)\gamma(F_{e^{h(x)}, \varepsilon}^h)\},$$

by (4.24) and (4.25) we obtain

$$(4.26) \quad \sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)} \leq \gamma(G).$$

Suppose that

$$\sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)} + \nu < \gamma(G)$$

for some $\nu > 0$. By taking $\lambda = 1$ and $0 < \varepsilon_0 < \nu/2$ in (4.22) we obtain

$$(4.27) \quad \gamma(F_{1, \varepsilon_0}^h) + \nu/2 < \gamma(G)$$

for all $h \in \mathcal{C}_{ba}(X)$ such that $e^h \leq 1_G$. Let (h_i) be an increasing net in $\mathcal{C}_{ba}(X)$ such that $\sup_i e^{h_i} = 1_G$, and let \wp be the set of finite subsets of I ordered by inclusion. Then $(\bigcup_{i \in \beta} G_{1, \varepsilon_0}^{h_i})_{\beta \in \wp}$ is an increasing net in \mathcal{G} such that

$$(4.28) \quad \forall \varepsilon > 0, \quad G \subset \bigcup_{\beta \in \wp} \bigcup_{i \in \beta} G_{1, \varepsilon_0}^{h_i}.$$

By (4.27) we have

$$\forall i \in I, \quad \gamma(\overline{G_{1, \varepsilon_0}^{h_i}}) + \nu/2 \leq \gamma(F_{1, \varepsilon_0}^{h_i}) + \nu/2 < \gamma(G),$$

and by (4.20),

$$\forall \beta \in \wp, \quad \gamma(\bigcup_{i \in \beta} G_{1, \varepsilon_0}^{h_i}) + \nu/2 < \gamma(G),$$

which contradicts (4.28) by (b). Therefore,

$$\sup_{h \in \mathcal{C}_{ba}(X), e^h \leq 1_G} e^{\Lambda(h)} = \gamma(G),$$

and by (4.4)

$$\sup_G e^{-J} = \gamma(G),$$

which gives (4.7) by upper semi-continuity of e^{-J} .

Suppose that (v) holds. By Theorem 3.3,

$$e^{\Lambda(h)} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \liminf \mu_\alpha^{t_\alpha}(G_{\lambda, \varepsilon}^h)\} = \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}^h)\}$$

for all $h \in \mathcal{C}_b(X)$. Let $(h_i)_{i \in I}$ be an increasing net in $\mathcal{C}_b(X)$ converging to $h \in \mathcal{C}_b(X)$. We will prove that $\lim \Lambda(h_i) = \Lambda(h)$. If $\Lambda(h) > \sup_{i \in I} \Lambda(h_i)$, then there exists $\lambda_0 \geq 0$, $\varepsilon_0 > 0$ and $\nu > 0$ such that

$$(4.29) \quad \begin{aligned} (\lambda_0 - \varepsilon_0) \liminf \mu_\alpha^{t_\alpha}(G_{\lambda_0, \varepsilon_0}^h) &> \sup_{i \in I} \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(F_{\lambda, \varepsilon}^{h_i})\} + \nu \\ &\geq \sup_{i \in I} \sup_{\lambda \geq 0, \varepsilon > 0} \{(\lambda - \varepsilon) \limsup \mu_\alpha^{t_\alpha}(\overline{G_{\lambda, \varepsilon}^{h_i}})\} + \nu \\ &\geq \sup_{i \in I} \sup_{\varepsilon > 0} \{(\lambda_0 - \varepsilon + \nu/2) \limsup \mu_\alpha^{t_\alpha}(\overline{G_{\lambda_0, \varepsilon}^{h_i}})\} + \nu/2. \end{aligned}$$

Take $\varepsilon_0 < \varepsilon < \varepsilon_0 + \nu/2$ in (4.29) and obtain

$$(4.30) \quad (\lambda_0 - \varepsilon_0) \liminf \mu_\alpha^{t_\alpha}(G_{\lambda_0, \varepsilon_0}^h) > (\lambda_0 - \varepsilon_0) \sup_{i \in I} \limsup \mu_\alpha^{t_\alpha}(\overline{G_{\lambda_0, \varepsilon}^{h_i}}) + \nu/2.$$

Put $F = F_{\lambda_0, \varepsilon_0}^h$, $G_i = G_{\lambda_0, \varepsilon}^{h_i}$ for all $i \in I$, and notice that $F \subset \bigcup_{i \in I} G_i$. Since $\overset{\circ}{F} \supset G_{\lambda_0, \varepsilon_0}^h$ we obtain by (4.30),

$$\liminf \mu_\alpha^{t_\alpha}(\overset{\circ}{F}) > \sup_{i \in I} \limsup \mu_\alpha^{t_\alpha}(\overline{G_i}) + \nu/2,$$

and so

$$\liminf \mu_{\alpha}^{t_{\alpha}}(\overset{\circ}{F}) > \limsup \mu_{\alpha}^{t_{\alpha}}\left(\bigcup_{1 \leq j \leq N} \overline{G_{i_j}}\right) + \nu/2,$$

for all finite subsets $\{G_{i_j} : 1 \leq j \leq N\} \subset \{G_i : i \in I\}$, which contradicts (4.1). Thus $\lim \Lambda(h_i) = \Lambda(h)$, that is, (ii) and so (i) hold.

It remains to prove $(i) \Leftrightarrow (vi)$, (4.8) and the last assertion. Suppose that (i) holds with rate function J . Put $f = e^{-J}$ and let $F \in \mathcal{F}$. By (4.3) we have

$$\sup_{x \in F} f(x) = \inf_{h \in \mathcal{C}_b(X), h|_F = 0} e^{\Lambda(h)} \leq \inf_{G \in \mathcal{G}, G \supset F} \liminf \mu_{\alpha}^{t_{\alpha}}(G) \leq \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G).$$

Suppose that

$$\inf_{h \in \mathcal{C}_b(X), h|_F = 0} e^{\Lambda(h)} < \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G).$$

Then, there exists $\nu > 0$ and $h_F \in \mathcal{C}_b(X)$ with $h_F|_F = 0$ such that

$$(4.31) \quad e^{\Lambda(h_F)} + \nu < \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G).$$

Since

$$(4.32) \quad e^{\Lambda(h_F)} = \sup_{\{x \in X; h_F(x) < \infty\}, \varepsilon > 0} \{(e^{h_F(x)} - \varepsilon) \limsup \mu_{\alpha}^{t_{\alpha}}(F_{e^{h_F(x)}, \varepsilon}^{h_F})\}$$

by Theorem 3.3, by taking $\varepsilon = \varepsilon_0 < \nu$ in (4.32) we obtain by (4.31),

$$(4.33) \quad \sup_{\{x \in X; h_F(x) < \infty\}} \{e^{h_F(x)} \limsup \mu_{\alpha}^{t_{\alpha}}(F_{e^{h_F(x)}, \varepsilon_0}^{h_F})\} < \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G).$$

Since $h_F(x) = 0$ for all $x \in F$, we have $F \subset G_{1, \varepsilon_0}^{h_F} \subset F_{1, \varepsilon_0}^{h_F}$ and (4.33) implies

$$\limsup \mu_{\alpha}^{t_{\alpha}}(G_{1, \varepsilon_0}^{h_F}) \leq \limsup \mu_{\alpha}^{t_{\alpha}}(F_{1, \varepsilon_0}^{h_F}) < \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G)$$

and the contradiction. We have shown that

$$\sup_{x \in F} f(x) = \inf_{G \in \mathcal{G}, G \supset F} \liminf \mu_{\alpha}^{t_{\alpha}}(G) = \inf_{G \in \mathcal{G}, G \supset F} \limsup \mu_{\alpha}^{t_{\alpha}}(G) \quad \text{for all } F \in \mathcal{F},$$

which is equivalent to (4.2) with $l = J$, and so (vi) holds.

If (vi) holds with $l : X \rightarrow [0, +\infty]$, then (4.2) implies

$$\limsup \mu_{\alpha}^{t_{\alpha}}(F) \leq \sup_F e^{-l} \leq \liminf \mu_{\alpha}^{t_{\alpha}}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$. By Proposition 2.3, (i) holds with rate function $\overset{\circ}{l}$ and (4.8) holds.

If moreover X is second countable, then it is well known that for any family $\{h_i : i \in I\}$ of lower semi-continuous functions on X there exists a countable subset $I_0 \subset I$ such that $\sup_{i \in I} h_i = \sup_{i \in I_0} h_i$. It is easy to see in the above proof that this property allows us to replace “net” by “sequence” in (ii) (resp. (iv)), and “open covers” by “countable open covers” in (v). \square

By combining Theorem 4.1 with Corollary 3.4 and Lemma 2.1, we obtain in the following corollary necessary and sufficient conditions in order that a large deviation principle occurs with rate function the lower regularization $\overset{\circ}{l}$ of a given function $l : X \rightarrow [0, +\infty]$. Notice that by Proposition 2.3 and $(i) \Leftrightarrow (ii)$ in Corollary 4.2, the infimum of the set of $[0, +\infty]$ -valued functions l on X satisfying (3.25) coincides with the lower regularization $\overset{\circ}{l}$ of each its elements. The equivalence $(i) \Leftrightarrow (iii)$ in

Corollary 4.2 was known when l is a tight rate function ([2], Theorem 4.4.13); here there is no hypothesis on l .

Corollary 4.2. *Suppose that X is normal, and let l be a $[0, +\infty]$ -valued function on X . Then, the following statements are equivalent:*

- (i) (μ_α) satisfies a large deviation principle with powers (t_α) and rate function $\overset{\circ}{l}$.
- (ii)

$$\limsup \mu_\alpha^{t_\alpha}(F) \leq \sup_{x \in G} e^{-l(x)} \leq \liminf \mu_\alpha^{t_\alpha}(G)$$

for all $F \in \mathcal{F}$, $G \in \mathcal{G}$ with $F \subset G$.

- (iii) $\Lambda(h)$ exists and

$$\Lambda(h) = \sup_{x \in X} \{h(x) - l(x)\} \quad \text{for all } h \in \mathcal{C}_b(X).$$

- (iv) $\Lambda(h)$ exists and

$$\Lambda(h) = \sup_{x \in X, h(x) < \infty} \{h(x) - l(x)\}$$

for all $[-\infty, +\infty]$ -valued continuous functions h on X satisfying (3.1).

Proof. (ii) \Rightarrow (iv) \Rightarrow (iii) by Corollary 3.4, (iii) \Rightarrow (i) by Theorem 4.1, and (i) \Rightarrow (ii) since $\sup_{x \in G} e^{-l(x)} = \sup_{x \in G} e^{-\overset{\circ}{l}(x)}$ for all $G \in \mathcal{G}$ by Lemma 2.1. \square

Recall that (μ_α) is said to be exponentially tight with respect to (t_α) if for all $\varepsilon > 0$ there is a compact set $K \subset X$ such that $\limsup \mu_\alpha^{t_\alpha}(X \setminus K) < \varepsilon$. Bryc's theorem ([2], Theorem 4.4.2) states that if $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$ and if (μ_α) is exponentially tight with respect to (t_α) , then (μ_α) satisfies a large deviation principle with powers (t_α) ; moreover, the (necessarily tight) rate function satisfies (4.5). The following Corollary 4.3 shows that the first conclusion holds under a hypothesis clearly weaker than exponential tightness. Moreover, Theorem 4.1 states that without any tightness hypothesis, a rate function for $(\mu_\alpha^{t_\alpha})$ always satisfies (4.5).

Corollary 4.3. *Suppose that X is normal. If $\Lambda(h)$ exists for all $h \in \mathcal{C}_b(X)$, and if for all open covers $\{G_i : i \in I\}$ of X , for all $\varepsilon > 0$, there exists a finite subset $\{G_{i_1}, \dots, G_{i_N}\} \subset \{G_i : i \in I\}$ such that*

$$(4.34) \quad \limsup \mu_\alpha^{t_\alpha}(X \setminus \bigcup_{1 \leq j \leq N} \overline{G_{i_j}}) < \varepsilon,$$

then (μ_α) satisfies a large deviation principle with powers (t_α) .

Proof. Let $F \in \mathcal{F}$, $\{G_i : i \in I\}$ be an open cover of F and $\varepsilon > 0$. Then, $\bigcup_{i \in I} G_i \cup X \setminus F$ is an open cover of X , and so there exists a finite subset $\{G_{i_1}, \dots, G_{i_N}\} \subset \{G_i : i \in I\}$ such that

$$\limsup \mu_\alpha^{t_\alpha}(X \setminus (\bigcup_{1 \leq j \leq N} \overline{G_{i_j}} \cup X \setminus \overset{\circ}{F})) < \varepsilon.$$

Thus,

$$\limsup \mu_\alpha^{t_\alpha}(\overset{\circ}{F} \setminus \bigcup_{1 \leq j \leq N} \overline{G_{i_j}}) < \varepsilon$$

and by Theorem 4.1, the conclusion holds. \square

Corollary 4.4. *If X is normal and $(\mu_\alpha^{t_\alpha})$ satisfies (4.34), then (μ_α) has a subnet (μ_β) satisfying a large deviation principle with powers (t_β) .*

Proof. Define $\Lambda_\alpha(h) = \log \mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$ for all $h \in C_b(X)$. Then, $(\Lambda_\alpha(\cdot))$ is a net in the compact space $[-\infty, +\infty]^{C_b(X)}$ (with the product topology), and so there is a subnet $(\Lambda_\beta(\cdot))$ converging to some limit $\Lambda'(\cdot)$. The result follows from Corollary 4.3 applied to $(\mu_\beta^{t_\beta})$. \square

Remark 4.5. $(i) \Leftrightarrow (iv)$ in Theorem 4.1 was known when γ is a sup-preserving capacity in the O'Brien sense, i.e., $\gamma(G) = \sup\{\gamma(K) : K \subset G, K \text{ compact}\}$ for all $G \in \mathcal{G}$, $\gamma(K) = \inf\{\gamma(G) : G \supset K, G \in \mathcal{G}\}$ for all compact $K \subset X$, and γ satisfies (2.1) of Lemma 2.1 ([3]). Thus, Theorem 4.1 removes the sup-preserving as well as the capacity conditions of γ (notice the difference between (iv) in Theorem 4.1 and (ii) in Proposition 2.3). In the spirit of Remark 3.5, this means that (μ_α) satisfies a large deviation principle with powers (t_α) if and only if $(\mu_\alpha^{t_\alpha})$ has a narrow set-theoretic limit in the set $\{\gamma \in \Gamma : \lim \gamma(G_i) = \gamma(\bigcup_i G_i) \text{ for all increasing nets } (G_i) \text{ in } \mathcal{G}\}$.

Remark 4.6. When X is Polish and $(\mu_\alpha) = (\mu_n)_{n \in \mathbf{N}_*}$, a recent result of Bryc and Bell ([1], Theorem 2.1) implies the equivalence of the following statements:

- (i') (μ_n) satisfies a large deviation principle with powers $(1/n)$ and tight rate function;
- (ii') $\Lambda(h)$ exists for all $h \in C_b(X)$, and $\inf_m \Lambda(h_m) = \Lambda(h)$ for each decreasing sequence (h_m) in $C_b(X)$ converging to $h \in C_b(X)$.

The equivalence $(i) \Leftrightarrow (ii)$ together with the last assertion in Theorem 4.1 can be seen as a free tightness analogue of that, by replacing "decreasing" by "increasing" in (ii'), and removing "tight" in (i').

Remark 4.7. The relation (4.2) in Theorem 4.1 generalizes a well-known expression of a rate function J for $(\mu_\alpha^{t_\alpha})$, obtained with $l = J$ and F ranging over all singletons ([2], Theorem 4.1.18).

REFERENCES

1. W. Bryc and H. Bell. Variational representations of Varadhan functionals, *Proc. Amer. Math. Soc.*, 129 (2001), No. 7, pp. 2119-2125. MR **2002b**:60040
2. A. Dembo and O. Zeitouni. Large deviations techniques and applications, *Second edition*, Springer-Verlag, New York, 1998. MR **99d**:60030
3. G. L. O'Brien and W. Verwaat. Capacities, large deviations and loglog laws. *Stable Processes and Related Topics (Ithaca, NY, 1990)*, pp. 43-83, *Progr. Probab.* 25, Birkhäuser, Boston, MA, 1991. MR **92k**:60007

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SANTIAGO OF CHILE, BERNARDO O'HIGGINS 3363, SANTIAGO, CHILE

E-mail address: hcomman@usach.cl

INTEGRATION BY PARTS FORMULAS INVOLVING GENERALIZED FOURIER-FEYNMAN TRANSFORMS ON FUNCTION SPACE

SEUNG JUN CHANG, JAE GIL CHOI, AND DAVID SKOUG

ABSTRACT. In an upcoming paper, Chang and Skoug used a generalized Brownian motion process to define a generalized analytic Feynman integral and a generalized analytic Fourier-Feynman transform. In this paper we establish several integration by parts formulas involving generalized Feynman integrals, generalized Fourier-Feynman transforms, and the first variation of functionals of the form $F(x) = f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$ where $\langle \alpha, x \rangle$ denotes the Paley-Wiener-Zygmund stochastic integral $\int_0^T \alpha(t) dx(t)$.

1. INTRODUCTION

In [11], Park and Skoug, working in the setting of one-parameter Wiener space $C_0[0, T]$ established several integration by parts formulas involving analytic Feynman integrals, Fourier-Feynman transforms, and the first variation of functionals of the form

$$(1.1) \quad F(x) = f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$$

where $\langle \alpha, x \rangle$ denotes the Paley-Wiener-Zygmund stochastic integral $\int_0^T \alpha(t) dx(t)$.

In this paper, we also study functionals of the form (1.1) but with x in a very general function space $C_{a,b}[0, T]$ rather than in the Wiener space $C_0[0, T]$. The Wiener process used in [11] is free of drift and is stationary in time while the stochastic process used in this paper is nonstationary in time, is subject to a drift $a(t)$, and can be used to explain the position of the Ornstein-Uhlenbeck process in an external force field [10]. It turns out, as noted in Remark 3.1 below, that including a drift term $a(t)$ makes establishing various integration by parts formulas for Fourier-Feynman transforms very difficult.

By choosing $a(t) = 0$ and $b(t) = t$ on $[0, T]$, the function space $C_{a,b}[0, T]$ reduces to the Wiener space $C_0[0, T]$, and so the results in [11] are immediate corollaries of the results in this paper. For related work see [3], [4], and [6].

Received by the editors September 6, 2002 and, in revised form, November 15, 2002.

2000 *Mathematics Subject Classification.* Primary 60J65, 28C20.

Key words and phrases. Generalized Brownian motion process, generalized analytic Feynman integral, generalized analytic Fourier-Feynman transform, first variation, Cameron-Storvick type theorem.

The present research was conducted by the research fund of Dankook University in 2000.

2. DEFINITIONS AND PRELIMINARIES

In this section we list the appropriate preliminaries and definitions from [5] that are needed to establish our parts formulas in Sections 3, 4 and 5 below.

Let $D = [0, T]$ and let (Ω, \mathcal{B}, P) be a probability measure space. A real-valued stochastic process Y on (Ω, \mathcal{B}, P) and D is called a *generalized Brownian motion process* if $Y(0, \omega) = 0$ almost everywhere and for $0 = t_0 < t_1 < \cdots < t_n \leq T$, the n -dimensional random vector $(Y(t_1, \omega), \dots, Y(t_n, \omega))$ is normally distributed with the density function

$$(2.1) \quad K(\vec{t}, \vec{\eta}) = ((2\pi)^n \prod_{j=1}^n (b(t_j) - b(t_{j-1})))^{-1/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{j=1}^n \frac{((\eta_j - a(t_j)) - (\eta_{j-1} - a(t_{j-1})))^2}{b(t_j) - b(t_{j-1})} \right\}$$

where $\vec{\eta} = (\eta_1, \dots, \eta_n)$, $\eta_0 = 0$, $\vec{t} = (t_1, \dots, t_n)$, $a(t)$ is an absolutely continuous real-valued function on $[0, T]$ with $a(0) = 0$, $a'(t) \in L^2[0, T]$, and $b(t)$ is a strictly increasing, continuously differentiable real-valued function with $b(0) = 0$ and $b'(t) > 0$ for each $t \in [0, T]$.

As explained in [13, pp. 18–20], Y induces a probability measure μ on the measurable space $(\mathbb{R}^D, \mathcal{B}^D)$ where \mathbb{R}^D is the space of all real-valued functions $x(t)$, $t \in D$, and \mathcal{B}^D is the smallest σ -algebra of subsets of \mathbb{R}^D with respect to which all the coordinate evaluation maps $e_t(x) = x(t)$ defined on \mathbb{R}^D are measurable. The triple $(\mathbb{R}^D, \mathcal{B}^D, \mu)$ is a probability measure space. This measure space is called the function space induced by the generalized Brownian motion process Y determined by $a(\cdot)$ and $b(\cdot)$.

We note that the generalized Brownian motion process Y determined by $a(\cdot)$ and $b(\cdot)$ is a Gaussian process with mean function $a(t)$ and covariance function $r(s, t) = \min\{b(s), b(t)\}$. By Theorem 14.2, [13, p. 187], the probability measure μ induced by Y , taking a separable version, is supported by $C_{a,b}[0, T]$ (which is equivalent to the Banach space of continuous functions x on $[0, T]$ with $x(0) = 0$ under the sup norm). Hence $(C_{a,b}[0, T], \mathcal{B}(C_{a,b}[0, T]), \mu)$ is the function space induced by Y where $\mathcal{B}(C_{a,b}[0, T])$ is the Borel σ -algebra of $C_{a,b}[0, T]$.

A subset B of $C_{a,b}[0, T]$ is said to be scale-invariant measurable [9] provided ρB is $\mathcal{B}(C_{a,b}[0, T])$ -measurable for all $\rho > 0$, and a scale-invariant measurable set N is said to be a scale-invariant null set provided $\mu(\rho N) = 0$ for all $\rho > 0$. A property that holds except on a scale-invariant null set is said to hold scale-invariant almost everywhere (s-a.e.).

Let $L^2_{a,b}[0, T]$ be the Hilbert space of functions on $[0, T]$ that are Lebesgue measurable and square integrable with respect to the Lebesgue-Stieltjes measures on $[0, T]$ induced by $a(\cdot)$ and $b(\cdot)$; i.e.,

$$(2.2) \quad L^2_{a,b}[0, T] = \left\{ v : \int_0^T v^2(s) db(s) < \infty \text{ and } \int_0^T v^2(s) d|a|(s) < \infty \right\}$$

where $|a|(t)$ denotes the total variation of the function a on the interval $[0, t]$.

For $u, v \in L^2_{a,b}[0, T]$, let

$$(2.3) \quad (u, v)_{a,b} = \int_0^T u(t)v(t) d[b(t) + |a|(t)].$$

Then $(\cdot, \cdot)_{a,b}$ is an inner product on $L^2_{a,b}[0, T]$ and $\|u\|_{a,b} = \sqrt{(u, u)_{a,b}}$ is a norm on $L^2_{a,b}[0, T]$. In particular, note that $\|u\|_{a,b} = 0$ if and only if $u(t) = 0$ a.e. on $[0, T]$. Furthermore, $(L^2_{a,b}[0, T], \|\cdot\|_{a,b})$ is a separable Hilbert space.

Let $\{\phi_j\}_{j=1}^\infty$ be a complete orthogonal set of real-valued functions of bounded variation on $[0, T]$ such that

$$(\phi_j, \phi_k)_{a,b} = \begin{cases} 0, & j \neq k, \\ 1, & j = k, \end{cases}$$

and for each $v \in L^2_{a,b}[0, T]$, let

$$(2.4) \quad v_n(t) = \sum_{j=1}^n (v, \phi_j)_{a,b} \phi_j(t)$$

for $n = 1, 2, \dots$. Then for each $v \in L^2_{a,b}[0, T]$, the Paley-Wiener-Zygmund (PWZ) stochastic integral $\langle v, x \rangle$ is defined by the formula

$$(2.5) \quad \langle v, x \rangle = \lim_{n \rightarrow \infty} \int_0^T v_n(t) dx(t)$$

for all $x \in C_{a,b}[0, T]$ for which the limit exists; one can show that for each $v \in L^2_{a,b}[0, T]$, the PWZ integral $\langle v, x \rangle$ exists for μ -a.e. $x \in C_{a,b}[0, T]$.

We denote the function space integral of a $\mathcal{B}(C_{a,b}[0, T])$ -measurable functional F by

$$(2.6) \quad E[F] = \int_{C_{a,b}[0, T]} F(x) d\mu(x)$$

whenever the integral exists.

We are now ready to state the definition of the generalized analytic Feynman integral.

Definition 2.1. Let \mathbb{C} denote the complex numbers. Let $\mathbb{C}_+ = \{\lambda \in \mathbb{C} : \operatorname{Re} \lambda > 0\}$ and $\tilde{\mathbb{C}}_+ = \{\lambda \in \mathbb{C} : \lambda \neq 0 \text{ and } \operatorname{Re} \lambda \geq 0\}$. Let $F : C_{a,b}[0, T] \rightarrow \mathbb{C}$ be such that for each $\lambda > 0$, the function space integral

$$J(\lambda) = \int_{C_{a,b}[0, T]} F(\lambda^{-\frac{1}{2}} x) d\mu(x)$$

exists for all $\lambda > 0$. If there exists a function $J^*(\lambda)$ analytic in \mathbb{C}_+ such that $J^*(\lambda) = J(\lambda)$ for all $\lambda > 0$, then $J^*(\lambda)$ is defined to be the analytic function space integral of F over $C_{a,b}[0, T]$ with parameter λ , and for $\lambda \in \mathbb{C}_+$ we write

$$(2.7) \quad E^{\text{an}\lambda}[F] \equiv E_x^{\text{an}\lambda}[F(x)] = J^*(\lambda).$$

Let $q \neq 0$ be a real number and let F be a functional such that $E^{\text{an}\lambda}[F]$ exists for all $\lambda \in \mathbb{C}_+$. If the following limit exists, we call it the generalized analytic Feynman integral of F with parameter q and we write

$$(2.8) \quad E^{\text{anf}_q}[F] \equiv E_x^{\text{anf}_q}[F(x)] = \lim_{\lambda \rightarrow -iq} E^{\text{an}\lambda}[F]$$

where λ approaches $-iq$ through values in \mathbb{C}_+ .

Next (see [5], [7], [1], [8], and [6]) we state the definition of the generalized analytic Fourier-Feynman transform (GFFT).

Definition 2.2. For $\lambda \in \mathbb{C}_+$ and $y \in C_{a,b}[0, T]$, let

$$(2.9) \qquad T_\lambda(F)(y) = E_x^{\text{an}\lambda}[F(y+x)].$$

For $p \in (1, 2]$, we define the L_p analytic GFFT, $T_q(p; F)$ of F , by the formula ($\lambda \in \mathbb{C}_+$),

$$(2.10) \qquad T_q(p; F)(y) = \lim_{\lambda \rightarrow -iq} T_\lambda(F)(y)$$

if it exists; i.e., for each $\rho > 0$,

$$\lim_{\lambda \rightarrow -iq} \int_{C_{a,b}[0, T]} |T_\lambda(F)(\rho y) - T_q(p; F)(\rho y)|^{p'} d\mu(y) = 0$$

where $1/p + 1/p' = 1$. We define the L_1 analytic GFFT, $T_q(1; F)$ of F , by the formula ($\lambda \in \mathbb{C}_+$)

$$(2.11) \qquad T_q(1; F)(y) = \lim_{\lambda \rightarrow -iq} T_\lambda(F)(y)$$

if it exists.

We note that for $1 \leq p \leq 2$, $T_q(p; F)$ is only defined as s-a.e. We also note that if $T_q(p; F)$ exists and if $F \approx G$, then $T_q(p; G)$ exists and $T_q(p; G) \approx T_q(p; F)$.

Next we give the definition of the first variation of a functional F on $C_{a,b}[0, T]$ followed by a very fundamental Cameron-Storvick type theorem [2] which was established in [5, Theorem 3.5].

Definition 2.3. Let F be a $\mathcal{B}(C_{a,b}[0, T])$ -measurable functional on $C_{a,b}[0, T]$ and let $w \in C_{a,b}[0, T]$. Then

$$(2.12) \qquad \delta F(x|w) = \left. \frac{\partial}{\partial h} F(x + hw) \right|_{h=0}$$

(if it exists) is called the first variation of F .

Throughout this paper, when working with $\delta F(x|w)$, we will always require w to be an element of A where

$$(2.13) \qquad A = \{w \in C_{a,b}[0, T] : w(t) = \int_0^t z(s)db(s) \text{ for some } z \in L^2_{a,b}[0, T]\}.$$

Note that for $F(x)$ of the form (1.1), $\delta F(x|w)$ acts like a directional derivative in the direction of w . For example, if $f(u_1, u_2) = \exp\{3u_1 + 4u_2\}$ and $F(x) = f(\langle \alpha_1, x \rangle, \langle \alpha_2, x \rangle)$, then

$$\begin{aligned} \delta F(x|w) &= [3\langle \alpha_1, w \rangle + 4\langle \alpha_2, w \rangle] \exp\{3\langle \alpha_1, x \rangle + 4\langle \alpha_2, x \rangle\} \\ &= \langle \alpha_1, w \rangle f_1(\langle \alpha_1, x \rangle, \langle \alpha_2, x \rangle) + \langle \alpha_2, w \rangle f_2(\langle \alpha_1, x \rangle, \langle \alpha_2, x \rangle). \end{aligned}$$

The following notation is used throughout the paper:

$$(2.14) \qquad (u, a') = \int_0^T u(t)a'(t)dt = \int_0^T u(t)da(t)$$

and

$$(2.15) \qquad (u^2, b') = \int_0^T u^2(t)b'(t)dt = \int_0^T u^2(t)db(t)$$

for $u \in L^2_{a,b}[0, T]$. Furthermore, for all $\lambda \in \tilde{\mathbb{C}}_+$, $\sqrt{\lambda}$ is always chosen to have positive real part.

Theorem 2.1. Let $z \in L^2_{a,b}[0, T]$ be given and for $t \in [0, T]$, let $w(t) = \int_0^t z(s)db(s)$. For each $\rho > 0$, let $F(\rho x)$ be μ -integrable on $C_{a,b}[0, T]$ and let $F(\rho x)$ have a first variation $\delta F(\rho x|\rho w)$ for all $x \in C_{a,b}[0, T]$ such that for some positive function $\eta(\rho)$,

$$\sup_{|h| \leq \eta(\rho)} |\delta F(\rho x + \rho h w|\rho w)|$$

is μ -integrable. Then if any two of the three generalized analytic Feynman integrals in the following equation exist, then the third one also exists, and equality holds:

$$(2.16) \quad E_x^{\text{anf}_q}[\delta F(x|w)] = -iqE_x^{\text{anf}_q}[F(x)\langle z, x \rangle] - (-iq)^{\frac{1}{2}}(z, a')E_x^{\text{anf}_q}[F(x)].$$

In fact, for each $\lambda \in \mathbb{C}_+$, the above conclusions also hold for analytic function space integrals,

$$(2.17) \quad E_x^{\text{an}\lambda}[\delta F(x|w)] = \lambda E_x^{\text{an}\lambda}[F(x)\langle z, x \rangle] - \sqrt{\lambda}(z, a')E_x^{\text{an}\lambda}[F(x)].$$

We finish this section by stating a very fundamental integration formula for the function space $C_{a,b}[0, T]$.

Let $\{\alpha_1, \dots, \alpha_n\}$ be an orthonormal set of functions from $(L^2_{a,b}[0, T], \|\cdot\|_{a,b})$, and for $j \in \{1, \dots, n\}$ let

$$(2.18) \quad A_j \equiv (\alpha_j, a') = \int_0^T \alpha_j(t)da(t)$$

and

$$(2.19) \quad B_j \equiv (\alpha_j^2, b') = \int_0^T \alpha_j^2(t)db(t).$$

Note that $B_j > 0$ for each $j \in \{1, 2, \dots, n\}$, while for each j , A_j may be positive, negative or zero.

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be Lebesgue measurable, and let $F(x) = f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$. Then

$$(2.20) \quad \begin{aligned} E[F] &\equiv \int_{C_{a,b}[0, T]} f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle) d\mu(x) \\ &= \left(\prod_{j=1}^n 2\pi B_j \right)^{-\frac{1}{2}} \int_{\mathbb{R}^n} f(u_1, \dots, u_n) \exp \left\{ - \sum_{j=1}^n \frac{(u_j - A_j)^2}{2B_j} \right\} du_1 \cdots du_n \end{aligned}$$

in the sense that if either side exists, both sides exist and equality holds.

3. INTEGRATION BY PARTS FORMULAS ON FUNCTION SPACE

Let n be a positive integer (fixed throughout this paper) and let $\{\alpha_1, \dots, \alpha_n\}$ be an orthonormal set of functions from $(L^2_{a,b}[0, T], \|\cdot\|_{a,b})$. Let m be a nonnegative integer. Then for $1 \leq p < \infty$, let $\mathcal{B}(p; m)$ be the space of all functionals of the form (1.1) for s-a.e. $x \in C_{a,b}[0, T]$ where all of the k th-order partial derivatives $f_{j_1, \dots, j_k}(u_1, \dots, u_n) = f_{j_1, \dots, j_k}(\vec{u})$ of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ are continuous and in $L^p(\mathbb{R}^n)$ for $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, \dots, n\}$. Also, let $\mathcal{B}(\infty; m)$ be the space of all functionals of the form (1.1) for s-a.e. $x \in C_{a,b}[0, T]$ where for $k = 0, 1, \dots, m$, all of the k th-order partial derivatives $f_{j_1, \dots, j_k}(\vec{u})$ of f are in $C_0(\mathbb{R}^n)$, the space of bounded continuous functions on \mathbb{R}^n that vanish at infinity.

Our first lemma follows directly from the definitions of $\delta F(x|w)$ and $\mathcal{B}(p; m)$.

Lemma 3.1. Let $1 \leq p \leq \infty$ be given, let m be a positive integer, let $F \in \mathcal{B}(p; m)$ be given by equation (1.1) and let w be an element of A . Then

$$(3.1) \quad \delta F(x|w) = \sum_{j=1}^n \langle \alpha_j, w \rangle f_j(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$$

for s -a.e. $x \in C_{a,b}[0, T]$. Furthermore, as a function of x , $\delta F(\cdot|w) \in \mathcal{B}(p; m-1)$.

Lemma 3.2. Let p, m and F be as in Lemma 3.1. Let $z \in L^2_{a,b}[0, T]$ be given, and for $t \in [0, T]$, let $w(t) = \int_0^t z(s) db(s)$. Let $G \in \mathcal{B}(p'; m)$ be given by

$$(3.2) \quad G(x) = g(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$$

for s -a.e. $x \in C_{a,b}[0, T]$. Define $R(x) = F(x)G(x)$ for $x \in C_{a,b}[0, T]$. Then $R \in \mathcal{B}(1; m)$, and as a function of x , $\delta R(\cdot|w) \in \mathcal{B}(1; m-1)$.

Proof. Let $r(u_1, \dots, u_n) = f(u_1, \dots, u_n)g(u_1, \dots, u_n)$. Then $R(x) = r(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$ is an element of $\mathcal{B}(1; m)$ since all of the k th-order partial derivatives of r are continuous and in $L^1(\mathbb{R})$ for $k = 0, 1, \dots, m$. Applying Lemma 3.1 we obtain that $\delta R(x|w)$, as a function of x , belongs to $\mathcal{B}(1; m-1)$. \square

Remark 3.1. Let F, G and R be as in Lemma 3.2 above. In evaluating $E[F(\lambda^{-\frac{1}{2}}x)]$, $E[R(\lambda^{-\frac{1}{2}}x)]$, and $E[\delta R(\lambda^{-\frac{1}{2}}x|w)]$ for $\lambda > 0$, the expression

$$(3.3) \quad H(\lambda; u_1, \dots, u_n) \equiv H(\lambda; \vec{u}) = \exp \left\{ - \sum_{j=1}^n \frac{(\sqrt{\lambda}u_j - A_j)^2}{2B_j} \right\}$$

occurs, where A_j and B_j are given by equations (2.18) and (2.19) above. Clearly, for $\lambda > 0$, $|H(\lambda; \vec{u})| \leq 1$ for all $\vec{u} \in \mathbb{R}^n$ since $B_j > 0$ for all $j = 1, \dots, n$. But for $\lambda \in \tilde{\mathbb{C}}_+$, $|H(\lambda; \vec{u})|$ is not necessarily bounded by 1. Note that for each $\lambda \in \tilde{\mathbb{C}}_+$, $\sqrt{\lambda} = c + id$ with $c \geq |d| \geq 0$. Hence, for each $\lambda \in \tilde{\mathbb{C}}_+$,

$$(3.4) \quad \begin{aligned} H(\lambda; \vec{u}) &= \exp \left\{ - \sum_{j=1}^n \frac{(\sqrt{\lambda}u_j - A_j)^2}{2B_j} \right\} \\ &= \exp \left\{ - \sum_{j=1}^n \frac{[(c^2 - d^2 + 2cdi)u_j^2 - 2(c + di)A_ju_j + A_j^2]}{2B_j} \right\}, \end{aligned}$$

and so

$$(3.5) \quad |H(\lambda; \vec{u})| = \exp \left\{ - \sum_{j=1}^n \frac{[(c^2 - d^2)u_j^2 - 2cA_ju_j + A_j^2]}{2B_j} \right\}.$$

Note that for $\lambda \in \mathbb{C}_+$, the case we consider throughout Section 3, $\operatorname{Re}(\sqrt{\lambda}) = c > |d| = |\operatorname{Im}(\sqrt{\lambda})| \geq 0$, which implies that $c^2 - d^2 > 0$. Hence, for each $\lambda \in \mathbb{C}_+$, $H(\lambda; \vec{u})$, as a function of \vec{u} , is an element of $L^p(\mathbb{R}^n)$ for all $p \in [1, +\infty]$; in fact, $H(\lambda; \vec{u})$ also belongs to $C_0(\mathbb{R}^n)$. These observations are critical to the development of the integration by parts formulas throughout Section 3.

In Sections 4 and 5 below we consider the case where $\lambda = -iq \in \tilde{\mathbb{C}}_+ - \mathbb{C}_+$. In this case, $\sqrt{\lambda} = \sqrt{-iq} = c + id$ with $c = \sqrt{|q|/2} = |d|$. Hence, for $\lambda = -iq$, $q \in \mathbb{R} - \{0\}$, $c^2 - d^2 = 0$, and so

$$(3.6) \quad |H(-iq; \vec{u})| = \exp \left\{ \sum_{j=1}^n \frac{[\sqrt{2|q|}A_ju_j - A_j^2]}{2B_j} \right\},$$

which is not necessarily in $L^p(\mathbb{R}^n)$ for any $p \in [1, +\infty]$. Thus, in Sections 4 and 5 we will need to put additional restrictions on the functionals F and G in order to obtain the corresponding parts formulas involving Fourier-Feynman transforms.

Remark 3.2. Note that in the setting of [11], $a(t) = 0$ and $b(t) = t$ on $[0, T]$ and so $A_j = (\alpha_j, a') = 0$ and $B_j = (\alpha_j^2, b') = 1$ for all $j \in \{1, 2, \dots, n\}$. Hence, for all $\lambda \in \tilde{\mathbb{C}}_+$,

$$|H(\lambda; \vec{u})| = \left| \exp \left\{ -\frac{\lambda}{2} \sum_{j=1}^n u_j^2 \right\} \right| = \exp \left\{ -\frac{\operatorname{Re}(\lambda)}{2} \sum_{j=1}^n u_j^2 \right\} \leq 1.$$

Theorem 3.3. Let $z \in L^2_{a,b}[0, T]$ be given and for $t \in [0, T]$, let $w(t) = \int_0^t z(s)db(s)$. Let p, m, F and G be as in Lemma 3.2. Then for all $\lambda \in \mathbf{C}_+$,

$$(3.7) \quad \begin{aligned} E_x^{\operatorname{an}\lambda} [F(x)\delta G(x|w) + \delta F(x|w)G(x)] \\ = \lambda E_x^{\operatorname{an}\lambda} [F(x)G(x)\langle z, x \rangle] - \sqrt{\lambda}(z, a') E_x^{\operatorname{an}\lambda} [F(x)G(x)] \end{aligned}$$

where $\sqrt{\lambda}$ is chosen to have positive real part.

Proof. First define $R(x) = F(x)G(x)$ and let

$$r(u_1, \dots, u_n) = f(u_1, \dots, u_n)g(u_1, \dots, u_n).$$

Then by Lemma 3.2, $R \in \mathcal{B}(1; m)$ and $\delta R(\cdot|w) \in \mathcal{B}(1; m-1)$. Furthermore, all of the k th-order partial derivatives of r are continuous and in $L^1(\mathbb{R}^n)$ for $k = 0, 1, \dots, m$. Hence, $R(\rho x)$ is μ -integrable on $C_{a,b}[0, T]$ for each $\rho > 0$. In addition, for s-a.e. $x \in C_{a,b}[0, T]$,

$$(3.8) \quad \begin{aligned} \delta R(x|w) &= F(x)\delta G(x|w) + \delta F(x|w)G(x) \\ &= f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle) \sum_{j=1}^n \langle \alpha_j, w \rangle g_j(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle) \\ &\quad + g(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle) \sum_{j=1}^n \langle \alpha_j, w \rangle f_j(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle). \end{aligned}$$

But for all $u \in L^2_{a,b}[0, T]$,

$$(3.9) \quad \begin{aligned} |\langle u, w \rangle| &= \left| \int_0^T u(s)dw(s) \right| \\ &= \left| \int_0^T u(s)z(s)db(s) \right| \\ &\leq \int_0^T |u(s)z(s)|d[b(s) + |a|(s)] \\ &= (|u|, |z|)_{a,b} \\ &\leq \|u\|_{a,b} \|z\|_{a,b}. \end{aligned}$$

In particular, since $\{\alpha_1, \dots, \alpha_n\}$ are orthonormal, $|\langle \alpha_j, w \rangle| \leq \|z\|_{a,b}$ for each $j \in \{1, 2, \dots, n\}$.

Next, using (3.8) and (3.9), we see that for $\rho > 0$ and $h > 0$,

$$\begin{aligned}
 & |\delta R(\rho x + \rho h w | \rho w)| \\
 & \leq \rho \|z\|_{a,b} |f(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)| \\
 & \quad \cdot \sum_{j=1}^n |g_j(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)| \\
 (3.10) \quad & + \rho \|z\|_{a,b} |g(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)| \\
 & \quad \cdot \sum_{j=1}^n |f_j(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)|.
 \end{aligned}$$

But this implies that $\delta R(\rho x + \rho h w | \rho w)$, as a function of x , is μ -integrable for all $\rho > 0$ and $h > 0$. This can be seen by integrating the right-hand side of (3.10) term by term. For example, using (2.20), we see that for any $l \in \{1, \dots, n\}$,

$$\begin{aligned}
 & E[|f(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)| \\
 & \quad \cdot |g_l(\langle \alpha_1, \rho x + \rho h w \rangle, \dots, \langle \alpha_n, \rho x + \rho h w \rangle)|] \\
 & = \left(\prod_{j=1}^n 2\pi \rho^2 B_j \right)^{-1/2} \int_{\mathbb{R}^n} |f(u_1, \dots, u_n) g_l(u_1, \dots, u_n)| \\
 (3.11) \quad & \cdot \exp \left\{ - \sum_{j=1}^n \frac{[u_j - \rho(A_j + h \langle \alpha_j, w \rangle)]^2}{2\rho^2 B_j} \right\} du_1 \cdots du_n \\
 & \leq \left(\prod_{j=1}^n 2\pi \rho^2 B_j \right)^{-1/2} \|f\|_p \|g_l\|_{p'} < \infty.
 \end{aligned}$$

Thus, using (3.10) and (3.11), we obtain that for $\rho > 0$ and $h > 0$,

$$\begin{aligned}
 & E[|\delta R(\rho x + \rho h w | \rho w)|] \\
 & \leq \rho \|z\|_{a,b} \left(\prod_{j=1}^n 2\pi \rho^2 B_j \right)^{-1/2} \left[\|f\|_p \sum_{l=1}^n \|g_l\|_{p'} + \|g\|_{p'} \sum_{l=1}^n \|f_l\|_p \right] < \infty.
 \end{aligned}$$

Next, using (3.8), (2.19), (3.3), and (3.4), we see that for all $\lambda > 0$,

$$\begin{aligned}
 & E[F(\lambda^{-\frac{1}{2}} x) \delta G(\lambda^{-\frac{1}{2}} x | w) + \delta F(\lambda^{-\frac{1}{2}} x | w) G(x)] \\
 (3.12) \quad & = \left(\prod_{j=1}^n \frac{\lambda}{2\pi B_j} \right)^{1/2} \int_{\mathbb{R}^n} [f(\vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle g_l(\vec{u}) \\
 & \quad + g(\vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle f_l(\vec{u})] H(\lambda; \vec{u}) d\vec{u}.
 \end{aligned}$$

But, as noted in Remark 3.1 above, for each $\lambda \in \mathbf{C}_+$, $H(\lambda; \vec{u})$ is an element of $C_0(\mathbb{R}^n)$, and so the integrand on the right-hand side of (3.12) is in $L^1(\mathbb{R}^n)$. Hence,

$$E_x^{\text{an}\lambda}[\delta R(x|w)] = E_x^{\text{an}\lambda}[F(x)\delta G(x|w) + \delta F(x|w)G(x)]$$

exists for all $\lambda \in \mathbf{C}_+$. A similar argument shows that the analytic function space integral $E_x^{\text{an}\lambda}[F(x)G(x)]$ also exists for all $\lambda \in \mathbf{C}_+$. Equation (3.7) now follows from Theorem 2.1 above; in particular, from equation (2.17) with $F(x)$ replaced with $R(x)$. \square

The following two corollaries are special cases of Theorem 3.3.

Corollary 3.4. *Let z, w , and m be as in Theorem 3.3. Let $F \in \mathcal{B}(2; m)$ be given by (1.1). Then for all $\lambda \in \mathbb{C}_+$,*

$$(3.13) \quad \begin{aligned} E_x^{\text{an}\lambda}[F(x)\delta F(x|w)] \\ = \frac{\lambda}{2} E_x^{\text{an}\lambda}[(F(x))^2\langle z, x \rangle] - \frac{\sqrt{\lambda}}{2}(z, a') E_x^{\text{an}\lambda}[(F(x))^2]. \end{aligned}$$

Proof. In Theorem 3.3, choose $p = 2$ and $G(x) = F(x)$. \square

Corollary 3.5. *Let z_1 and z_2 be elements of $L_{a,b}^2[0, T]$, and for $t \in [0, T]$, let $w_j(t) = \int_0^t z_j(s)db(s)$ for $j \in \{1, 2\}$. Let $F \in \mathcal{B}(2; m)$ be given by equation (1.1). Then for all $\lambda \in \mathbb{C}_+$,*

$$(3.14) \quad \begin{aligned} E_x^{\text{an}\lambda}[F(x)\delta^2 F(\cdot|w_1)(x|w_2) + \delta F(x|w_2)\delta F(x|w_1)] \\ = \lambda E_x^{\text{an}\lambda}[F(x)\delta F(x|w_1)\langle z_2, x \rangle] - \sqrt{\lambda}(z_2, a') E_x^{\text{an}\lambda}[F(x)\delta F(x|w_1)]. \end{aligned}$$

Proof. Let $p = 2$ and $G(x) = \delta F(x|w_1)$ in Theorem 3.3. \square

Lemma 3.6. *Let p, m and F be as in Lemma 3.1 above. Then for all $\lambda \in \mathbb{C}_+$,*

$$(3.15) \quad T_\lambda(F)(y) = E_x^{\text{an}\lambda}[F(y+x)] = \phi_0(\lambda; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle)$$

for s-a.e. $y \in C_{a,b}[0, T]$ where

$$(3.16) \quad \phi_0(\lambda; \xi_1, \dots, \xi_n) = \left(\prod_{j=1}^n \frac{\lambda}{2\pi B_j} \right)^{1/2} \int_{\mathbb{R}^n} f(\vec{u} + \vec{\xi}) H(\lambda; \vec{u}) d\vec{u}$$

with B_j and H given by equations (2.19) and (3.4) respectively.

Proof. For $\lambda > 0$, equation (3.15) follows easily from equation (2.20). But for each $\lambda \in \mathbb{C}_+$, as shown in Remark 3.1 above, $H(\lambda; u_1, \dots, u_n)$ is an element of $L^p(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$ for all $p \in [1, \infty]$. Hence, for each $\lambda \in \mathbb{C}_+$ and s-a.e. $y \in C_{a,b}[0, T]$,

$$f(u_1 + \langle \alpha_1, y \rangle, \dots, u_n + \langle \alpha_n, y \rangle) H(\lambda; u_1, \dots, u_n)$$

belongs to $L^1(\mathbb{R}^n)$ and so equation (3.15) holds throughout \mathbb{C}_+ . \square

Our next lemma follows from standard results for convolution products. The key is that for each $\lambda \in \mathbb{C}_+$, $H(\lambda; \vec{u})$ is an element of $L^p(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$ for all $1 \leq p \leq +\infty$.

Lemma 3.7. *Let ϕ_0 be given by equation (3.16) above.*

(a) *If $f \in L^1(\mathbb{R}^n)$, then $\phi_0(\lambda; \cdot) \in C_0(\mathbb{R}^n)$ for all $\lambda \in \mathbb{C}_+$.*

(b) *If $f \in L^p(\mathbb{R}^n)$ for some $p \in (1, \infty)$, then $\phi_0(\lambda; \cdot) \in L^{p'}(\mathbb{R}^n)$ for all $\lambda \in \mathbb{C}_+$ where $p' = \frac{p}{p-1}$.*

(c) *If $f \in C_0(\mathbb{R}^n)$, then $\phi_0(\lambda; \cdot) \in L^1(\mathbb{R}^n)$ for all $\lambda \in \mathbb{C}_+$.*

Our next theorem follows immediately from Lemma 3.7.

Theorem 3.8. *Let $1 \leq p \leq \infty$ be given. If $F \in \mathcal{B}(p; m)$, then $T_\lambda(F) \in \mathcal{B}(p'; m)$.*

Theorem 3.9. Let $1 \leq p \leq \infty$ and $w \in A$ be given. Let $F \in \mathcal{B}(p; m)$ be given by equation (1.1). Then for all $\lambda \in \mathbb{C}_+$ and s-a.e. $y \in C_{a,b}[0, T]$,

$$\begin{aligned}
 & \delta T_\lambda(F)(y|w) \\
 (3.17) \quad &= \left(\prod_{j=1}^n \frac{\lambda}{2\pi B_j} \right)^{1/2} \int_{\mathbb{R}^n} \sum_{l=1}^n \langle \alpha_l, w \rangle f_l(u_1 + \langle \alpha_1, y \rangle, \dots, u_n + \langle \alpha_n, y \rangle) \\
 & \quad \cdot H(\lambda; u_1, \dots, u_n) du_1 \cdots du_n \\
 &= T_\lambda(\delta F(\cdot|w))(y),
 \end{aligned}$$

which, as a function of y , is an element of $\mathcal{B}(p'; m-1)$.

Proof. The fact that $\delta T_\lambda(F)(y|w)$ is an element of $\mathcal{B}(p'; m-1)$ follows directly from Theorem 3.8 and Lemma 3.1. To establish equation (3.17) for $\lambda > 0$, simply calculate $\delta T_\lambda(F)(y|w)$ using equation (3.15), and then calculate $T_\lambda(\delta F(\cdot|w))(y)$ using equations (3.1) and (2.9). Finally, equation (3.17) holds throughout \mathbb{C}_+ by analytic continuation in λ . \square

In our next theorem we obtain an integration by parts formula involving $T_\lambda(F)$ and $T_\lambda(G)$.

Theorem 3.10. Let p, m, z, w, F and G be as in Theorem 3.3. Then for all $\lambda \in \mathbb{C}_+$,

$$\begin{aligned}
 (3.18) \quad & E_x^{\text{an}\lambda} [T_\lambda(F)(x) \delta T_\lambda(G)(x|w) + \delta T_\lambda(F)(x|w) T_\lambda(G)(x)] \\
 &= \lambda E_\lambda^{\text{an}\lambda} [T_\lambda(F)(x) T_\lambda(G)(x) \langle z, x \rangle] - \sqrt{\lambda}(z, a') E_x^{\text{an}\lambda} [T_\lambda(F)(x) T_\lambda(G)(x)].
 \end{aligned}$$

Proof. For $x \in C_{a,b}[0, T]$, let $R(x) = T_\lambda(F)(x) T_\lambda(G)(x)$. Then by Theorem 3.8, $T_\lambda(F) \in \mathcal{B}(p'; m)$ and $T_\lambda(G) \in \mathcal{B}(p; m)$. Hence, by Lemma 3.2, R belongs to $\mathcal{B}(1; m)$, and so by Lemma 3.1, $\delta R(x|w)$, as a function of x , belongs to $\mathcal{B}(1; m-1)$. Thus, equation (3.18) follows from Theorem 3.3 with F and G replaced by $T_\lambda(F)$ and $T_\lambda(G)$ respectively. \square

Theorem 3.11. Let m, z and w be as in Lemma 3.2. Let $p \in [1, 2]$ and let F and G in $\mathcal{B}(p; m)$ be given by equations (1.1) and (3.2) respectively. Then for all $\lambda \in \mathbb{C}_+$,

$$\begin{aligned}
 (3.19) \quad & E_x^{\text{an}\lambda} [F(x) \delta T_\lambda(G)(x|w) + \delta F(x|w) T_\lambda(G)(x)] \\
 &= \lambda E_x^{\text{an}\lambda} [F(x) T_\lambda(G)(x) \langle z, x \rangle] - \sqrt{\lambda}(z, a') E_x^{\text{an}\lambda} [F(x) T_\lambda(G)(x)].
 \end{aligned}$$

Proof. Let $R(x) = F(x) T_\lambda(G)(x)$ for $x \in C_{a,b}[0, T]$. By Theorem 3.8, $T_\lambda(G)$ is an element of $\mathcal{B}(p'; m)$ and hence by Lemma 3.2, R belongs to $\mathcal{B}(1; m)$. Hence, by Lemma 3.1, $\delta R(x|w)$, as a function of x , belongs to $\mathcal{B}(1; m-1)$. Thus, equation (3.19) follows from Theorem 3.3 with G replaced by $T_\lambda(G)$. \square

Corollary 3.12. Let m, z, w, p and F be as in Theorem 3.11. Then for all $\lambda \in \mathbb{C}_+$,

$$\begin{aligned}
 (3.20) \quad & E_x^{\text{an}\lambda} [F(x) \delta T_\lambda(F)(x|w) + \delta F(x|w) T_\lambda(F)(x)] \\
 &= \lambda E_x^{\text{an}\lambda} [F(x) T_\lambda(F)(x) \langle z, x \rangle] - \sqrt{\lambda}(z, a') E_x^{\text{an}\lambda} [F(x) T_\lambda(F)(x)].
 \end{aligned}$$

Proof. Simply choose $G = F$ in Theorem 3.11. \square

Corollary 3.13. *Let m, z and w be as in Lemma 3.2. Let $F \in \mathcal{B}(2; m)$ be given by equation (1.1). Then for all $\lambda \in \mathbb{C}_+$,*

$$(3.21) \quad \begin{aligned} & E_x^{\text{an}\lambda} [T_\lambda(F)(x) \delta T_\lambda(F)(x|w)] \\ &= \frac{\lambda}{2} E_x^{\text{an}\lambda} [(T_\lambda(F)(x))^2 \langle z, x \rangle] - \frac{\sqrt{\lambda}(z, a')}{2} E_x^{\text{an}\lambda} [(T_\lambda(F)(x))^2]. \end{aligned}$$

Proof. Simply choose $p = 2$ and $G = F$ in Theorem 3.10. \square

4. PARTS FORMULAS INVOLVING $T_q(1; F)$ AND $T_q(1; G)$

In this section we obtain various integration by parts formulas involving the analytic GFFTs $T_q(1; F)$ and $T_q(1; G)$. In view of equation (3.6) above, we clearly need to impose additional restrictions on the functionals F and G than were needed throughout Section 3.

Fix $q \in \mathbb{R} - \{0\}$. Then as $\lambda \rightarrow -iq$ through values in \mathbb{C}_+ , $c = \text{Re}(\sqrt{\lambda}) \rightarrow \sqrt{|q|/2}$ and $|d| \rightarrow \sqrt{|q|/2}$ where $d = \text{Im}(\sqrt{\lambda})$.

Next using equations (3.3) through (3.6) we see that for all $\lambda \in \tilde{\mathbb{C}}_+$ with $c = \text{Re}(\sqrt{\lambda}) < ((1 + |q|)/2)^{\frac{1}{2}}$,

$$(4.1) \quad \begin{aligned} |H(\lambda; \vec{u})| &= \exp \left\{ - \sum_{j=1}^n \frac{[(c^2 - d^2)u_j^2 - 2cA_j u_j + A_j^2]}{2B_j} \right\} \\ &\leq \exp \left\{ \sum_{j=1}^n \frac{cA_j u_j}{B_j} \right\} \\ &\leq \exp \left\{ \left(\frac{1 + |q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\}. \end{aligned}$$

In addition,

$$(4.2) \quad \begin{aligned} & \int_{\mathbb{R}^n} |f(\vec{\xi} + \vec{u}) H(\lambda; \vec{u})| d\vec{u} \\ &= \int_{\mathbb{R}^n} |f(\vec{u}) H(\lambda; \vec{u} - \vec{\xi})| d\vec{u} \\ &\leq \exp \left\{ - \text{Re}(\sqrt{\lambda}) \sum_{j=1}^n \frac{A_j \xi_j}{B_j} \right\} \\ &\quad \cdot \int_{\mathbb{R}^n} |f(\vec{u})| \exp \left\{ \left(\frac{1 + |q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u}. \end{aligned}$$

For $f \in L^1(\mathbb{R}^n)$ let

$$(4.3) \quad \mathcal{F}(f)(\vec{\xi}) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(\vec{u}) \exp \left\{ i \sum_{j=1}^n u_j \xi_j \right\} d\vec{u}$$

denote the Fourier transform of f .

Theorem 4.1. *Let $q \in \mathbb{R} - \{0\}$ be given. Let $F \in \mathcal{B}(1; m)$ be given by equation (1.1) with*

$$(4.4) \quad \int_{\mathbb{R}^n} |f_{j_1, \dots, j_k}(\vec{u})| \exp \left\{ \left(\frac{1 + |q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} < \infty$$

for all $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, \dots, n\}$. Furthermore, assume that

$$(4.5) \qquad \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2|q|} A_j \xi_j}{2B_j} \right\} \mathcal{F}(f(\cdot)H(-iq; \cdot)) \left(- \frac{q\xi_1}{B_1}, \dots, - \frac{q\xi_n}{B_n} \right)$$

belongs to $C_0(\mathbb{R}^n)$. Then

$$(4.6) \qquad \phi_0(-iq; \vec{\xi}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u}$$

is an element of $C_0(\mathbb{R}^n)$. Furthermore, the L_1 analytic GFFT, $T_q(1; F)$ exists as an element of $\mathcal{B}(\infty; m)$ and for s-a.e. $y \in C_{a,b}[0, T]$ is given by the formula

$$(4.7) \qquad T_q(1; F)(y) = \phi_0(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle).$$

Proof. By (4.1) and (4.4) we know that $f(\cdot)H(-iq; \cdot) \in L^1(\mathbb{R}^n)$, and so its Fourier transform, $\mathcal{F}(f(\cdot)H(-iq; \cdot))(\vec{\xi})$ exists and belongs to $C_0(\mathbb{R}^n)$. Furthermore, by equations (4.6) and (3.4) and the fact that $\sqrt{-iq} = c + di = \sqrt{|q|}/2 + di$, we obtain

$$\begin{aligned} &\phi_0(-iq; \vec{\xi}) \\ &= \left(\prod_{j=1}^n -\frac{iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u} \\ &= \left(\prod_{j=1}^n -\frac{iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{\xi} + \vec{u}) \exp \left\{ - \sum_{j=1}^n \frac{[\sqrt{-iq}u_j - A_j]^2}{2B_j} \right\} d\vec{u} \\ &= \left(\prod_{j=1}^n -\frac{iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{u}) \exp \left\{ - \sum_{j=1}^n \frac{[\sqrt{-iq}(u_j - \xi_j) - A_j]^2}{2B_j} \right\} d\vec{u} \\ &= \left(\prod_{j=1}^n -\frac{iq}{B_j} \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[2\sqrt{-iq}A_j\xi_j - iq\xi_j^2]}{2B_j} \right\} \\ (4.8) \qquad &\cdot (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(\vec{u}) H(-iq; \vec{u}) \exp \left\{ - iq \sum_{j=1}^n \frac{u_j \xi_j}{B_j} \right\} d\vec{u} \\ &= \left(\prod_{j=1}^n -\frac{iq}{B_j} \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[2\sqrt{-iq}A_j\xi_j - iq\xi_j^2]}{2B_j} \right\} \\ &\qquad \cdot \mathcal{F}(f(\cdot)H(-iq; \cdot)) \left(- \frac{q\xi_1}{B_1}, \dots, - \frac{q\xi_n}{B_n} \right) \\ &= \left(\prod_{j=1}^n -\frac{iq}{B_j} \right)^{\frac{1}{2}} \exp \left\{ i \sum_{j=1}^n \frac{[q\xi_j^2 - 2dA_j\xi_j]}{2B_j} \right\} \\ &\qquad \cdot \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2|q|} A_j \xi_j}{2B_j} \right\} \mathcal{F}(f(\cdot)H(-iq; \cdot)) \left(- \frac{q\xi_1}{B_1}, \dots, - \frac{q\xi_n}{B_n} \right). \end{aligned}$$

By assumption (4.5), it follows that $\phi_0(-iq; \vec{\xi})$ is an element of $C_0(\mathbb{R}^n)$.

Finally, by equations (2.11), (3.15), (3.16), (4.8) and the dominated convergence theorem (the use of which is justified by (4.2)), it follows that for s-a.e. $y \in$

$$C_{a,b}[0, T],$$

$$\begin{aligned}
 & T_q(1; F)(y) \\
 &= \lim_{\lambda \rightarrow -iq} T_\lambda(F)(y) \\
 &= \lim_{\lambda \rightarrow -iq} \phi_0(\lambda; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle) \\
 (4.9) \quad &= \lim_{\lambda \rightarrow -iq} \left(\prod_{j=1}^n \frac{\lambda}{2\pi B_j} \right)^{1/2} \int_{\mathbb{R}^n} f(u_1 + \langle \alpha_1, y \rangle, \dots, u_n + \langle \alpha_n, y \rangle) H(\lambda; \vec{u}) d\vec{u} \\
 &= \left(\prod_{j=1}^n -\frac{iq}{2\pi B_j} \right)^{1/2} \int_{\mathbb{R}^n} f(u_1 + \langle \alpha_1, y \rangle, \dots, u_n + \langle \alpha_n, y \rangle) H(-iq; \vec{u}) d\vec{u} \\
 &= \phi_0(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle)
 \end{aligned}$$

as desired. \square

Theorem 4.2. Let $q \in \mathbb{R} - \{0\}$ and $F \in \mathcal{B}(1; m)$ be as in Theorem 4.1. Furthermore, assume that for each $l \in \{1, 2, \dots, n\}$,

$$(4.10) \quad \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2|q|} A_j \xi_j}{2B_j} \right\} \mathcal{F}(f_l(\cdot) H(-iq; \cdot)) \left(-\frac{q\xi_1}{B_1}, \dots, -\frac{q\xi_n}{B_n} \right)$$

belongs to $C_0(\mathbb{R}^n)$. Then for each $l \in \{1, 2, \dots, n\}$,

$$(4.11) \quad \phi_l(-iq; \vec{\xi}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f_l(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u}$$

is an element of $C_0(\mathbb{R}^n)$. In addition, for each $w \in A$ and s-a.e. $y \in C_{a,b}[0, T]$,

$$\begin{aligned}
 (4.12) \quad \delta T_q(1; F)(y|w) &= \sum_{l=1}^n \langle \alpha_l, w \rangle \phi_l(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle) \\
 &= T_q(1; \delta F(\cdot|w))(y),
 \end{aligned}$$

which, as a function of y , is an element of $\mathcal{B}(\infty; m-1)$.

Proof. The proof that each $\phi_l(-iq; \cdot)$ belongs to $C_0(\mathbb{R}^n)$ is the same as the proof in Theorem 4.1 above showing that $\phi_0(-iq; \cdot) \in C_0(\mathbb{R}^n)$. Equation (4.12) then follows immediately using the definition of the first variation and equation (4.7). \square

Our next theorem gives a parts formula involving F and $T_q(1; G)$.

Theorem 4.3. Let $q \in \mathbb{R} - \{0\}$ be given and let $F \in \mathcal{B}(1; m)$ be as in Theorem 4.1. Let $G \in \mathcal{B}(1; m)$ be given by equation (3.2) with

$$(4.13) \quad \int_{\mathbb{R}^n} |g_{j_1, \dots, j_k}(\vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} < \infty$$

for all $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, \dots, n\}$. Furthermore, assume that

$$(4.14) \quad \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2|q|} A_j \xi_j}{2B_j} \right\} \mathcal{F}(g_l(\cdot) H(-iq; \cdot)) \left(-\frac{q\xi_1}{B_1}, \dots, -\frac{q\xi_n}{B_n} \right)$$

belongs to $C_0(\mathbb{R}^n)$ for all $l \in \{0, 1, \dots, n\}$. Let $z \in L^2_{a,b}[0, T]$ be given and for $t \in [0, T]$, let $w(t) = \int_0^t z(s)db(s)$. Then

(4.15)

$$\begin{aligned} E_x^{\text{anf}_q}[F(x)\delta T_q(1; G)(x|w) + \delta F(x|w)T_q(1; G)(x)] \\ = -iqE_x^{\text{anf}_q}[F(x)T_q(1; G)(x)\langle z, x \rangle] \\ - (-iq)^{\frac{1}{2}}(z, a')E_x^{\text{anf}_q}[F(x)T_q(1; G)(x)]. \end{aligned}$$

Proof. Let $R(x) = F(x)T_q(1; G)(x)$. By Theorem 4.1, $T_q(1; G)(x)$ is an element of $\mathcal{B}(\infty; m)$ and so $R(x)$ is an element of $\mathcal{B}(1; m)$. Also, by Theorem 4.1, Theorem 4.2 and Lemma 3.2,

$$\delta R(x|w) = F(x)\delta T_q(1; G)(x|w) + \delta F(x|w)T_q(1; G)(x),$$

as a function of x , is an element of $\mathcal{B}(1; m - 1)$. In addition, we know that for each $l \in \{0, 1, \dots, n\}$,

$$\psi_l(-iq; \vec{v}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j}\right)^{1/2} \int_{\mathbb{R}^n} g_l(\vec{u} + \vec{v})H(-iq; \vec{u})d\vec{u}$$

is an element of $C_0(\mathbb{R}^n)$ with

$$T_q(1; G)(y) = \psi_0(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle)$$

and

$$\delta T_q(1; G)(y|w) = \sum_{l=1}^n \langle \alpha_l, w \rangle \psi_l(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle)$$

for s-a.e. $y \in C_{a,b}[0, T]$. Hence, both of the following analytic Feynman integrals exist:

(4.16)

$$\begin{aligned} E_x^{\text{anf}_q}[R(x)] &= E_x^{\text{anf}_q}[F(x)T_q(1; G)(x)] \\ &= \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j}\right)^{1/2} \int_{\mathbb{R}^n} f(\vec{u})\psi_0(-iq; \vec{u})H(-iq; \vec{u})d\vec{u} \end{aligned}$$

and

(4.17)

$$\begin{aligned} E_x^{\text{anf}_q}[\delta R(x|w)] &= E_x^{\text{anf}_q}[F(x)\delta T_q(1; G)(x|w) + \delta F(x|w)T_q(1; G)(x)] \\ &= \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j}\right)^{1/2} \int_{\mathbb{R}^n} \left[f(\vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle \psi_l(-iq; \vec{u}) \right. \\ &\quad \left. + \psi_0(-iq; \vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle f_l(\vec{u}) \right] H(-iq; \vec{u})d\vec{u}. \end{aligned}$$

Also, proceeding as in the proof of Theorem 3.3 above, it is easy to show that for $\rho > 0$ and $h > 0$,

(4.18)

$$\begin{aligned} E[|\delta R(\rho x + \rho h w|\rho w)|] \\ \leq \rho \|z\|_{a,b} \left(\prod_{j=1}^n 2\pi \rho^2 B_j\right)^{-1/2} \left[\|f(\cdot)\|_1 \sum_{l=1}^n \|\psi_l(-iq; \cdot)\|_\infty \right. \\ \left. + \|\psi_0(-iq; \cdot)\|_\infty \sum_{l=1}^n \|f_l(\cdot)\|_1 \right] < \infty. \end{aligned}$$

Hence, by Theorem 2.1 above, the analytic Feynman integral

$$E_x^{\text{anf}_q}[R(x)\langle z, x \rangle] = E_x^{\text{anf}_q}[F(x)T_q(1; G)(x)\langle z, x \rangle]$$

exists and equality (4.14) holds. \square

Choosing $G = F$ in Theorem 4.3 we get the following integration by parts formula.

Corollary 4.4. *Let $q \in \mathbb{R} - \{0\}$ be given and let $F \in \mathcal{B}(1; m)$ be as in Theorem 4.2. Let z and w be as in Theorem 4.3. Then*

$$\begin{aligned} E_x^{\text{anf}_q}[F(x)\delta T_q(1; F)(x|w) + \delta F(x|w)T_q(1; F)(x)] \\ (4.19) \quad = -iqE_x^{\text{anf}_q}[F(x)T_q(1; F)(x)\langle z, x \rangle] \\ - (-iq)^{\frac{1}{2}}(z, a')E_x^{\text{anf}_q}[F(x)T_q(1; F)(x)]. \end{aligned}$$

Next we obtain a parts formula involving $T_q(1; F)$ and $T_q(1; G)$.

Theorem 4.5. *Let $q \in \mathbb{R} - \{0\}$. Let $F \in \mathcal{B}(1; m)$ be as in Theorem 4.2 and let $G \in \mathcal{B}(1; m)$ be as in Theorem 4.3. Furthermore, assume that for each $l \in \{0, 1, \dots, n\}$,*

$$(4.20) \quad \int_{\mathbb{R}^n} |\psi_l(-iq; \vec{u})H(-iq; \vec{u})| d\vec{u} < \infty.$$

Then for $w(t) = \int_0^t z(s)db(s)$ with $z \in L_{a,b}^2[0, T]$,

$$\begin{aligned} E_x^{\text{anf}_q}[T_q(1; F)(x)\delta T_q(1; G)(x|w) + \delta T_q(1; F)(x|w)T_q(1; G)(x)] \\ (4.21) \quad = -iqE_x^{\text{anf}_q}[T_q(1; F)(x)T_q(1; G)(x)\langle z, x \rangle] \\ - (-iq)^{\frac{1}{2}}(z, a')E_x^{\text{anf}_q}[T_q(1; F)(x)T_q(1; G)(x)]. \end{aligned}$$

Proof. Let $R(x) = T_q(1; F)(x)T_q(1; G)(x)$. Then $R \in \mathcal{B}(\infty; m)$ and $\delta R(x|w)$, as a function of x , is an element of $\mathcal{B}(\infty; m-1)$. Hence, by (4.6), (4.11) and (4.20), both of the following analytic Feynman integrals exist:

$$(4.22) \quad E_x^{\text{anf}_q}[R(x)] = \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} \phi_0(-iq; \vec{u})\psi_0(-iq; \vec{u})H(-iq; \vec{u})d\vec{u}$$

and

$$\begin{aligned} E_x^{\text{anf}_q}[\delta R(x|w)] \\ (4.23) \quad = \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} \left[\phi_0(-iq; \vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle \psi_l(-iq; \vec{u}) \right. \\ \left. + \psi_0(-iq; \vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle \phi_l(-iq; \vec{u}) \right] H(-iq; \vec{u})d\vec{u}. \end{aligned}$$

In addition, for $\rho > 0$ and $h > 0$,

$$\begin{aligned} E[|\delta R(\rho x + \rho h w|\rho w)|] \\ (4.24) \quad \leq \rho \|z\|_{a,b} \left[\|\phi_0(-iq; \cdot)\|_{\infty} \sum_{l=1}^n \|\psi_l(-iq; \cdot)\|_{\infty} \right. \\ \left. + \|\psi_0(-iq; \cdot)\|_{\infty} \sum_{l=1}^n \|\phi_l(-iq; \cdot)\|_{\infty} \right] < \infty. \end{aligned}$$

Hence, by Theorem 2.1, the analytic Feynman integral $E_x^{\text{anf}_q}[R(x)\langle z, x \rangle]$ exists and equality (4.21) holds. \square

We finish this section with some examples which shed light upon the necessity of conditions such as (4.4) and (4.5), and which also illustrate that the conclusions of Lemma 3.7 are not necessarily valid for $\lambda \in \tilde{\mathbb{C}}_+$ with $\text{Re}(\lambda) = 0$.

In our first example we define a functional F of the form (1.1) with $n = 1$, such that F is an element of $\mathcal{B}(p; m)$ for all $p \in [1, +\infty]$, f is an element of $L^p(\mathbb{R})$ for all $p \in [1, +\infty]$, and yet $\phi_0(i; \cdot)$ given by (4.6) is not an element of $C_0(\mathbb{R})$. In fact, $|\phi_0(i; \xi_1)| = +\infty$ for all $\xi_1 \in \mathbb{R}$.

Example 4.6. Let $q = -1$, let $n = 1$, let m be a nonnegative integer, and let α_1 be an element of $L^2_{a,b}[0, T]$ with $\|\alpha_1\|_{a,b} = 1$. Without loss of generality, we will assume that A_1 (see equation (2.18)) is positive.

Let $f : \mathbb{R} \rightarrow \mathbb{C}$ be defined by the formula

$$(4.25) \quad f(u_1) \equiv u_1^{m+1} \chi_{[0, +\infty)}(u_1) \exp \left\{ \frac{i u_1^2}{2B_1} - \frac{i \sqrt{2} A_1 u_1}{2B_1} + \frac{A_1^2}{2B_1} - \frac{\sqrt{2} A_1 u_1}{4B_1} \right\}.$$

We note that

$$|f(u_1)| = u_1^{m+1} \chi_{[0, +\infty)}(u_1) \exp \left\{ \frac{A_1^2}{2B_1} - \frac{\sqrt{2} A_1 u_1}{4B_1} \right\},$$

and hence $f \in L^p(\mathbb{R})$ for all $p \in [1, +\infty]$. In fact, f is also an element of $C_0(\mathbb{R})$. We then define $F : C_{a,b}[0, T] \rightarrow \mathbb{C}$ by the formula

$$(4.27) \quad F(x) \equiv f(\langle \alpha_1, x \rangle).$$

It is easy to see that $F \in \mathcal{B}(p; m)$ for all $p \in [1, +\infty]$.

Next, using equation (3.4) with $n = 1$, $\lambda = i$, and $\sqrt{i} = \frac{1+i}{\sqrt{2}}$, we observe that

$$(4.28) \quad H(i; u_1) = \exp \left\{ \frac{\sqrt{2} A_1 u_1 + i \sqrt{2} A_1 u_1 - A_1^2 - i u_1^2}{2B_1} \right\},$$

and hence

$$(4.29) \quad f(u_1) H(i; u_1) = u_1^{m+1} \chi_{[0, +\infty)}(u_1) \exp \left\{ \frac{\sqrt{2} A_1 u_1}{4B_1} \right\},$$

which is not an element of $L^p(\mathbb{R})$ for any $p \in [1, +\infty]$.

Then, using equation (4.6) with $n = 1$ and $q = -1$, equation (4.25) and equation (4.28), we see that

$$\begin{aligned} \phi_0(i; \xi_1) &= \left(\frac{i}{2\pi B_1} \right)^{\frac{1}{2}} \int_{\mathbb{R}} f(u_1 + \xi_1) H(i; u_1) du_1 \\ &= \left(\frac{i}{2\pi B_1} \right)^{\frac{1}{2}} \exp \left\{ \frac{i \xi_1^2}{2B_1} - \frac{i \sqrt{2} A_1 \xi_1}{2B_1} - \frac{\sqrt{2} A_1 \xi_1}{4B_1} \right\} \\ (4.30) \quad &\cdot \int_{\mathbb{R}} (u_1 + \xi_1)^{m+1} \chi_{[0, +\infty)}(u_1 + \xi_1) \exp \left\{ \frac{i u_1 \xi_1}{B_1} + \frac{\sqrt{2} A_1 u_1}{4B_1} \right\} du_1 \\ &= \left(\frac{i}{2\pi B_1} \right)^{\frac{1}{2}} \exp \left\{ \frac{i \xi_1^2}{2B_1} - \frac{i \sqrt{2} A_1 \xi_1}{2B_1} - \frac{\sqrt{2} A_1 \xi_1}{4B_1} \right\} \\ &\cdot \int_{-\xi_1}^{+\infty} (u_1 + \xi_1)^{m+1} \exp \left\{ \frac{i u_1 \xi_1}{B_1} + \frac{\sqrt{2} A_1 u_1}{4B_1} \right\} du_1. \end{aligned}$$

Thus,

$$(4.31) \quad |\phi_0(i; \xi_1)| = (2\pi B_1)^{-\frac{1}{2}} \exp \left\{ -\frac{\sqrt{2}A_1\xi_1}{4B_1} \right\} \cdot \left| \int_{-\xi_1}^{+\infty} (u_1 + \xi_1)^{m+1} \exp \left\{ \frac{iu_1\xi_1}{B_1} + \frac{\sqrt{2}A_1u_1}{4B_1} \right\} du_1 \right|.$$

Hence, choosing $\xi_1 = 0$, and using the fact that A_1 is positive, we see that

$$\begin{aligned} |\phi_0(i; 0)| &= (2\pi B_1)^{-\frac{1}{2}} \left| \int_0^{+\infty} u^{m+1} \exp \left\{ \frac{\sqrt{2}A_1u}{4B_1} \right\} du \right| \\ &= (2\pi B_1)^{-\frac{1}{2}} \int_0^{+\infty} u^{m+1} \exp \left\{ \frac{\sqrt{2}A_1u}{4B_1} \right\} du = +\infty, \end{aligned}$$

which implies that $\phi_0(i; \cdot)$ is not an element of $C_0(\mathbb{R})$. In fact, for each fixed $\xi_1 \in \mathbb{R}$, we observe that

$$(4.32) \quad |\phi_0(i; \xi_1)| = (2\pi B_1)^{-\frac{1}{2}} \exp \left\{ -\frac{\sqrt{2}A_1\xi_1}{4B_1} \right\} \cdot \left| \int_{-\xi_1}^{+\infty} (u_1 + \xi_1)^{m+1} \exp \left\{ \frac{iu_1\xi_1}{B_1} + \frac{\sqrt{2}A_1u_1}{4B_1} \right\} du_1 \right| = +\infty,$$

and so $\phi_0(i; \cdot)$ is not an element of $L^p(\mathbb{R})$ for any $p \in [1, +\infty]$ even though $f(\cdot)$ was an element of $L^p(\mathbb{R})$ for all $p \in [1, +\infty]$ and F was an element of $\mathcal{B}(p; m)$ for all $p \in [1, +\infty]$. Hence, the L_1 analytic GFFT, $T_{-1}(1; F)$ does not exist.

We also note that f does not satisfy condition (4.4) above since by equation (4.26) (recall that $q = -1$ and so $(\frac{1+|q|}{2})^{1/2} = 1$),

$$\begin{aligned} &\int_{\mathbb{R}} |f(u_1)| \exp \left\{ \frac{|A_1u_1|}{B_1} \right\} du_1 \\ &= \int_0^{+\infty} u_1^{m+1} \exp \left\{ \frac{A_1^2}{2B_1} - \frac{\sqrt{2}A_1u_1}{4B_1} + \frac{A_1u_1}{B_1} \right\} du_1 = +\infty. \end{aligned}$$

In our next example we exhibit a functional F of the form (1.1) that satisfies conditions (4.4) and (4.5) above. Furthermore, we are able to evaluate the integral in equation (4.6) and thus obtain a formula for $\phi_0(i; \vec{\xi})$ which does not involve any integrals.

Example 4.7. Let $q = -1$, let m be a nonnegative integer and let n be a positive integer. Let $\{\alpha_1, \dots, \alpha_n\}$ be an orthonormal set of functions from $(L_{a,b}^2[0, T], \|\cdot\|_{a,b})$, and for each $j \in \{1, \dots, n\}$ let A_j and B_j be given by (2.18) and (2.19) respectively. We define $f: \mathbb{R}^n \rightarrow \mathbb{C}$ by the formula

$$(4.33) \quad f(\vec{u}) \equiv \exp \left\{ \sum_{j=1}^n \frac{[iu_j^2 - i\sqrt{2}A_ju_j + A_j^2 - u_j^2 - \sqrt{2}A_ju_j]}{2B_j} \right\}.$$

We note that

$$(4.34) \quad |f(\vec{u})| = \exp \left\{ \sum_{j=1}^n \frac{[A_j^2 - u_j^2 - \sqrt{2}A_ju_j]}{2B_j} \right\},$$

and hence $f \in L^p(\mathbb{R}^n)$ for all $p \in [1, +\infty]$. Also, $f \in C_0(\mathbb{R}^n)$.

Let $F : C_{a,b}[0, T] \rightarrow \mathbb{C}$ be given by

$$(4.35) \quad F(x) \equiv f(\langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle).$$

It is easy to show that $F \in \mathcal{B}(p; m)$ for all $p \in [1, +\infty]$.

Next, using equation (4.33), together with equation (3.4) with $\lambda = i$ and $\sqrt{i} = \frac{1+i}{\sqrt{2}}$, it follows that

$$(4.36) \quad f(\vec{u})H(i; \vec{u}) = \exp \left\{ - \sum_{j=1}^n \frac{u_j^2}{2B_j} \right\}.$$

Now clearly $f(\cdot)H(i; \cdot)$ is an element of $L^p(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$ for all $p \in [1, +\infty]$. Next, using equations (4.6), (3.4) and (4.33) we obtain

$$\begin{aligned} \phi_0(i; \vec{\xi}) &= \left(\prod_{j=1}^n \frac{i}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{u} + \vec{\xi})H(i; \vec{u})d\vec{u} \\ &= \left(\prod_{j=1}^n \frac{i}{2\pi B_j} \right)^{\frac{1}{2}} \exp \left\{ \sum_{j=1}^n \frac{[i\xi_j^2 - i\sqrt{2}A_j\xi_j - \sqrt{2}A_j\xi_j]}{2B_j} \right\} \\ &\quad \cdot \int_{\mathbb{R}^n} \exp \left\{ i \sum_{j=1}^n \frac{u_j\xi_j}{B_j} - \sum_{j=1}^n \frac{(u_j + \xi_j)^2}{2B_j} \right\} d\vec{u} \\ (4.37) \quad &= \left(\prod_{j=1}^n \frac{i}{2\pi B_j} \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[i\xi_j^2 + i\sqrt{2}A_j\xi_j + \sqrt{2}A_j\xi_j]}{2B_j} \right\} \\ &\quad \cdot \int_{\mathbb{R}^n} \exp \left\{ i \sum_{j=1}^n \frac{u_j\xi_j}{B_j} - \sum_{j=1}^n \frac{u_j^2}{2B_j} \right\} d\vec{u} \\ &= (i)^{\frac{n}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[i\xi_j^2 + i\sqrt{2}A_j\xi_j + \sqrt{2}A_j\xi_j + \xi_j^2]}{2B_j} \right\}, \end{aligned}$$

because

$$\begin{aligned} &\int_{\mathbb{R}^n} \exp \left\{ i \sum_{j=1}^n \frac{u_j\xi_j}{B_j} - \sum_{j=1}^n \frac{u_j^2}{2B_j} \right\} d\vec{u} \\ (4.38) \quad &= \left(\prod_{j=1}^n 2\pi B_j \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{\xi_j^2}{2B_j} \right\} \\ &= (2\pi)^{\frac{n}{2}} \mathcal{F}(f(\cdot)H(i; \cdot)) \left(\frac{\xi_1}{B_1}, \dots, \frac{\xi_n}{B_n} \right). \end{aligned}$$

Hence,

$$(4.39) \quad |\phi_0(i; \vec{\xi})| = \exp \left\{ - \sum_{j=1}^n \frac{[\sqrt{2}A_j\xi_j + \xi_j^2]}{2B_j} \right\},$$

and so $\phi_0(i; \cdot)$ is an element of $C_0(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ for all $p \in [1, +\infty]$.

We also note that because of the factor $\exp\{-\frac{u_j^2}{2B_j}\}$ in the definition of $f(\vec{u})$ given by equation (4.33), condition (4.4) certainly holds. In addition, condition

(4.5) holds because (recall $q = -1$)

$$\begin{aligned}
 & \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2} A_j \xi_j}{2 B_j} \right\} \mathcal{F}(f(\cdot) H(i; \cdot)) \left(\frac{\xi_1}{B_1}, \dots, \frac{\xi_n}{B_n} \right) \\
 &= \exp \left\{ - \sum_{j=1}^n \frac{\sqrt{2} A_j \xi_j}{2 B_j} \right\} \left(\prod_{j=1}^n B_j \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{\xi_j^2}{2 B_j} \right\} \\
 (4.40) \quad &= \left(\prod_{j=1}^n B_j \right)^{\frac{1}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[\sqrt{2} A_j \xi_j + \xi_j^2]}{2 B_j} \right\} \\
 &= \left(\prod_{j=1}^n B_j \right)^{\frac{1}{2}} |\phi_0(i; \vec{\xi})|
 \end{aligned}$$

is an element of $C_0(\mathbb{R}^n)$ as was shown above. Hence, by Theorem 4.1, the L_1 analytic GFFT, $T_{-1}(1; F)$ exists as an element of $\mathcal{B}(\infty; m)$ and for s-a.e. $y \in C_{a,b}[0, T]$ is given by the formula

$$\begin{aligned}
 & T_{-1}(1; F)(y) \\
 &= \phi_0(i; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle) \\
 (4.41) \quad &= (i)^{\frac{n}{2}} \exp \left\{ - \sum_{j=1}^n \frac{[(1+i)\langle \alpha_j, y \rangle^2 + (1+i)\sqrt{2} A_j \langle \alpha_j, y \rangle]}{2 B_j} \right\}.
 \end{aligned}$$

5. PARTS FORMULAS INVOLVING $T_q(2; F)$ AND $T_q(2; G)$

Note that in our first theorem below we replace conditions (4.4) and (4.5) with condition (5.1). This condition is used to obtain a dominating function in order to apply the dominated convergence theorem.

Theorem 5.1. *Let $q \in \mathbb{R} - \{0\}$ be given. Let $F \in \mathcal{B}(2; m)$ be given by equation (1.1) with*

$$(5.1) \quad \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |f_{j_1, \dots, j_k}(\vec{\xi} + \vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} \right]^2 d\vec{\xi} < \infty$$

for all $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, \dots, n\}$. Then

$$(5.2) \quad \phi_0(-iq; \vec{\xi}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u}$$

is an element of $L^2(\mathbb{R}^n)$. Furthermore, the L_2 analytic GFFT, $T_q(2; F)$ exists as an element of $\mathcal{B}(2; m)$ and for s-a.e. $y \in C_{a,b}[0, T]$ is given by the formula

$$(5.3) \quad T_q(2; F)(y) = \phi_0(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle).$$

Proof. Using (4.1) we first note that

$$\begin{aligned}
 |\phi_0(-iq; \vec{\xi})| &\leq \left(\prod_{j=1}^n \frac{|q|}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} |f(\vec{\xi} + \vec{u}) H(-iq; \vec{u})| d\vec{u} \\
 &\leq \left(\prod_{j=1}^n \frac{|q|}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} |f(\vec{\xi} + \vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u}.
 \end{aligned}$$

Hence, by (5.1) with $k = 0$,

$$\begin{aligned} \|\phi_0(-iq; \cdot)\|_2^2 &= \int_{\mathbb{R}^n} |\phi_0(-iq; \vec{\xi})|^2 d\vec{\xi} \\ &\leq \left(\prod_{j=1}^n \frac{|q|}{2\pi B_j} \right) \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |f(\vec{\xi} + \vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} \right]^2 d\vec{\xi} < \infty \end{aligned}$$

and so $\phi_0(-iq; \vec{\xi})$ is an element of $L^2(\mathbb{R}^n)$.

To show that $T_q(2; F)$ exists and is given by equation (5.3) it suffices to show that for each $\rho > 0$,

$$\lim_{\lambda \rightarrow -iq} \int_{C_{a,b}[0,T]} |T_\lambda(\rho y) - \phi_0(-iq; \langle \alpha_1, \rho y \rangle, \dots, \langle \alpha_n, \rho y \rangle)|^2 d\mu(y) = 0.$$

But

$$\begin{aligned} &\int_{C_{a,b}[0,T]} |T_\lambda(\rho y) - \phi_0(-iq; \langle \alpha_1, \rho y \rangle, \dots, \langle \alpha_n, \rho y \rangle)|^2 d\mu(y) \\ &= \int_{C_{a,b}[0,T]} |\phi_0(\lambda; \langle \alpha_1, \rho y \rangle, \dots, \langle \alpha_n, \rho y \rangle) \\ &\quad - \phi_0(-iq; \langle \alpha_1, \rho y \rangle, \dots, \langle \alpha_n, \rho y \rangle)|^2 d\mu(y) \\ &= \left(\prod_{j=1}^n 2\pi B_j \rho^2 \right)^{-\frac{1}{2}} \int_{\mathbb{R}^n} |\phi_0(\lambda; \vec{u}) - \phi_0(-iq; \vec{u})|^2 \exp \left\{ - \sum_{j=1}^n \frac{(u_j - \rho A_j)^2}{2\rho^2 B_j} \right\} d\vec{u} \\ &\leq \left(\prod_{j=1}^n 2\pi B_j \rho^2 \right)^{-\frac{1}{2}} \|\phi_0(\lambda; \cdot) - \phi_0(-iq; \cdot)\|_2^2. \end{aligned}$$

Now clearly $\phi_0(\lambda; \vec{\xi}) \rightarrow \phi_0(-iq; \vec{\xi})$ a.e. on \mathbb{R}^n as $\lambda \rightarrow -iq$ through values in \mathbb{C}_+ . Thus, to show that

$$\|\phi_0(\lambda; \cdot) - \phi_0(-iq; \cdot)\|_2 \rightarrow 0,$$

it suffices [11, p. 126] to show that

$$\|\phi_0(\lambda; \cdot)\|_2 \rightarrow \|\phi_0(-iq; \cdot)\|_2$$

as $\lambda \rightarrow -iq$ through values in \mathbb{C}_+ . But for all $\lambda \in \mathbb{C}_+$ with $\text{Re}(\sqrt{\lambda}) < ((1+|q|)/2)^{\frac{1}{2}}$,

$$\begin{aligned} &\int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} f(\vec{\xi} + \vec{u}) H(\lambda; \vec{u}) d\vec{u} \right|^2 d\vec{\xi} \\ &\leq \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |f(\vec{\xi} + \vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} \right]^2 d\vec{\xi} < \infty. \end{aligned}$$

Hence, by the dominated convergence theorem,

$$\begin{aligned}
 & \lim_{\lambda \rightarrow -iq} \|\phi_0(\lambda; \cdot)\|_2^2 \\
 &= \lim_{\lambda \rightarrow -iq} \int_{\mathbb{R}^n} \left| \left(\prod_{j=1}^n \frac{\lambda}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{u} + \vec{\xi}) H(\lambda; \vec{u}) d\vec{u} \right|^2 d\vec{\xi} \\
 &= \lim_{\lambda \rightarrow -iq} \left(\prod_{j=1}^n \frac{|\lambda|}{2\pi B_j} \right) \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} f(\vec{u} + \vec{\xi}) H(\lambda; \vec{u}) d\vec{u} \right|^2 d\vec{\xi} \\
 &= \left(\prod_{j=1}^n \frac{|q|}{2\pi B_j} \right) \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} f(\vec{u} + \vec{\xi}) H(-iq; \vec{u}) d\vec{u} \right|^2 d\vec{\xi} \\
 &= \|\phi_0(-iq; \cdot)\|_2^2.
 \end{aligned}$$

□

Corollary 5.2. *Let $q \in \mathbb{R} - \{0\}$ and $F \in \mathcal{B}(2; m)$ be as in Theorem 5.1. Then for each $l \in \{1, 2, \dots, n\}$,*

$$(5.4) \quad \phi_l(-iq; \vec{\xi}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f_l(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u}$$

is an element of $L^2(\mathbb{R}^n)$. In addition, for each $w \in A$ and s-a.e. $y \in C_{a,b}[0, T]$,

$$\begin{aligned}
 (5.5) \quad \delta T_q(2; F)(y|w) &= \sum_{l=1}^n \langle \alpha_l, w \rangle \phi_l(-iq; \langle \alpha_1, y \rangle, \dots, \langle \alpha_n, y \rangle) \\
 &= T_q(2; \delta F(\cdot|w))(y),
 \end{aligned}$$

which, as a function of y , is an element of $\mathcal{B}(2; m-1)$.

Proof. The proof that each $\phi_l(-iq; \cdot)$ belongs to $L^2(\mathbb{R}^n)$ is the same as the proof in Theorem 5.1 above showing that $\phi_0(-iq, \cdot) \in L^2(\mathbb{R}^n)$. Equation (5.5) then follows immediately using the definition of the first variation and equations (5.2) and (5.3). □

Theorem 5.3. *Let $q \in \mathbb{R} - \{0\}$ be given and let $F \in \mathcal{B}(2; m)$ be as in Theorem 5.1. Furthermore, assume that*

$$(5.6) \quad \int_{\mathbb{R}^n} |f_{j_1, \dots, j_k}(\vec{u}) H(-iq; \vec{u})|^2 d\vec{u} < \infty$$

for all $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, 2, \dots, n\}$.

Let $G \in \mathcal{B}(2; m)$ be given by equation (3.2) with

$$(5.7) \quad \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |g_{j_1, \dots, j_k}(\vec{\xi} + \vec{u})| \exp \left\{ \left(\frac{1+|q|}{2} \right)^{\frac{1}{2}} \sum_{j=1}^n \frac{|A_j u_j|}{B_j} \right\} d\vec{u} \right]^2 d\vec{\xi} < \infty$$

for all $k \in \{0, 1, \dots, m\}$ and each $j_i \in \{1, \dots, n\}$. Let z and w be as in Theorem 4.3. Then

$$\begin{aligned}
 (5.8) \quad & E_x^{\text{anf}_q} [F(x) \delta T_q(2; G)(x|w) + \delta F(x|w) T_q(2; G)(x)] \\
 &= -iq E_x^{\text{anf}_q} [F(x) T_q(2; G)(x) \langle z, x \rangle] \\
 &\quad - (-iq)^{\frac{1}{2}} (z, a') E_x^{\text{anf}_q} [F(x) T_q(2; G)(x)].
 \end{aligned}$$

Proof. By Theorem 5.1, for each $l \in \{0, 1, \dots, n\}$,

(5.9)
$$\psi_l(-iq; \vec{\xi}) \equiv \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} g_l(\vec{\xi} + \vec{u}) H(-iq; \vec{u}) d\vec{u}$$

is an element of $L^2(\mathbb{R}^n)$. Furthermore,

(5.10)
$$T_q(2; G)(x) = \psi_0(-iq; \langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$$

is an element of $\mathcal{B}(2; m)$, and as a function of x ,

(5.11)
$$\delta T_q(2; G)(x|w) = \sum_{l=1}^n \langle \alpha_l, w \rangle \psi_l(-iq; \langle \alpha_1, x \rangle, \dots, \langle \alpha_n, x \rangle)$$

belongs to $\mathcal{B}(2; m - 1)$. Hence, $R(x) = F(x)T_q(2; G)(x)$ is an element of $\mathcal{B}(1; m)$ and

$$\delta R(x|w) = F(x)\delta T_q(2; G)(x|w) + \delta F(x|w)T_q(2; G)(x)$$

is an element of $\mathcal{B}(1; m - 1)$. Since

$$f(\vec{u})H(-iq; \vec{u})\psi_l(-iq; \vec{u}) \quad \text{and} \quad f_l(\vec{u})H(-iq; \vec{u})\psi_0(-iq; \vec{u})$$

belong to $L^1(\mathbb{R}^n)$ for each $l \in \{0, 1, \dots, n\}$, both of the following analytic Feynman integrals exist:

$$E_x^{\text{anf}_q}[R(x)] = \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} f(\vec{u})H(-iq; \vec{u})\psi_0(-iq; \vec{u})d\vec{u}$$

and

$$\begin{aligned} E_x^{\text{anf}_q}[\delta R(x|w)] &= \left(\prod_{j=1}^n \frac{-iq}{2\pi B_j} \right)^{\frac{1}{2}} \int_{\mathbb{R}^n} \left[f(\vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle \psi_l(-iq; \vec{u}) \right. \\ &\quad \left. + \psi_0(-iq; \vec{u}) \sum_{l=1}^n \langle \alpha_l, w \rangle f_l(\vec{u}) \right] H(-iq; \vec{u})d\vec{u}. \end{aligned}$$

Also, for $\rho > 0$ and $h > 0$,

$$\begin{aligned} &E[|\delta R(\rho x + \rho h w|\rho w)|] \\ &\leq \rho \|z\|_{a,b} \left(\prod_{j=1}^n 2\pi \rho^2 B_j \right)^{-\frac{1}{2}} \left[\|f\|_2 \sum_{l=1}^n \|\psi_l(-iq; \cdot)\|_2 \right. \\ &\quad \left. + \|\psi_0(-iq; \cdot)\|_2 \sum_{l=1}^n \|f_l\|_2 \right] < \infty. \end{aligned}$$

Hence, by Theorem 2.1, the analytic Feynman integral $E_x^{\text{anf}_q}[R(x)\langle z, x \rangle]$ exists and equality (5.8) holds. □

Next, choosing $G = F$ in Theorem 5.3, we obtain the following integration by parts formula.

Corollary 5.4. *Let $q, F \in \mathcal{B}(2; m)$, z , and w be as in Theorem 5.3. Then*

$$\begin{aligned} &E_x^{\text{anf}_q}[F(x)\delta T_q(2; F)(x|w) + \delta F(x|w)T_q(2; F)(x)] \\ &= -iqE_x^{\text{anf}_q}[F(x)T_q(2; F)(x)\langle z, x \rangle] \\ &\quad - (-iq)^{\frac{1}{2}}(z, a')E_x^{\text{anf}_q}[F(x)T_q(2; F)(x)]. \end{aligned}$$

Our final theorem is a counterpart to Theorem 4.5 above.

Theorem 5.5. *Let $q \in \mathbb{R} - \{0\}$ and let F and G be as in Theorem 5.3. Furthermore, assume that for each $l \in \{0, 1, \dots, n\}$,*

$$(5.12) \quad \int_{\mathbb{R}^n} |\psi_l(\vec{u}) H(-iq; \vec{u})|^2 d\vec{u} < \infty.$$

Let $z \in L_{a,b}^2[0, T]$ be given and for $t \in [0, T]$ let $w(t) = \int_0^t z(s) db(s)$. Then

$$(5.13) \quad \begin{aligned} & E_x^{\text{anf}_q} [T_q(2; F)(x) \delta T_q(2; G)(x|w) + \delta T_q(2; F)(x|w) T_q(2; G)(x)] \\ &= -iq E_x^{\text{anf}_q} [T_q(2; F)(x) T_q(2; G)(x) \langle z, x \rangle] \\ &\quad - (-iq)^{\frac{1}{2}} (z, a') E_x^{\text{anf}_q} [T_q(2; F)(x) T_q(2; G)(x)]. \end{aligned}$$

Proof. Let $R(x) = T_q(2; F)(x) T_q(2; G)(x)$. Then $R \in \mathcal{B}(1; m)$ and $\delta R(\cdot|w) \in \mathcal{B}(1; m-1)$. Using (5.2), (5.4), (5.9) and (5.12), we see that $E_x^{\text{anf}_q} [R(x)]$ and $E_x^{\text{anf}_q} [\delta R(x|w)]$ both exist and are given by equations (4.22) and (4.23) respectively. Finally, we see that (5.13) follows from Theorem 2.1, since for $\rho > 0$ and $h > 0$,

$$\begin{aligned} & E[|\delta R(\rho x + \rho h w|\rho w)|] \\ &\leq \rho \|z\|_{a,b} \left(\prod_{j=1}^n 2\pi \rho^2 B_j \right)^{-\frac{1}{2}} [\|\phi_0(-iq; \cdot)\|_2 \sum_{l=1}^n \|\psi_l(-iq; \cdot)\|_2 \\ &\quad + \|\psi_0(-iq; \cdot)\|_2 \sum_{l=1}^n \|\phi_l(-iq; \cdot)\|_2] < \infty. \end{aligned}$$

□

We finish this paper with some very brief comments about the functionals defined in Examples 4.6 and 4.7 for the case $p = 2$.

We first note that for the functional $F(x) \in \mathcal{B}(2; m)$ defined by equation (4.27) with $f(u_1) \in L^2(\mathbb{R})$ given by (4.25), the L_2 analytic GFFT, $T_{-1}(2; F)$ does not exist because $|\phi_0(i; \xi_1)| = +\infty$ for each $\xi_1 \in \mathbb{R}$. In fact, the L_p analytic GFFT, $T_{-1}(p; F)$ does not exist for any $p \in [1, 2]$.

On the other hand, it is quite easy to see that condition (5.1) holds for the function $f(\vec{u})$ given by equation (4.33). Hence, for $F(x)$ defined by equation (4.35), the L_2 analytic GFFT, $T_{-1}(2; F)$ exists as an element of $\mathcal{B}(2; m)$ and for s.a.e. $y \in C_{a,b}[0, T]$ is given by the right-hand side of equation (4.41). In fact, for all $p \in [1, 2]$, the L_p analytic GFFT, $T_{-1}(p; F)$ exists as an element of $\mathcal{B}(p'; m)$ and is given by the right-hand side of equation (4.41).

REFERENCES

- [1] R. H. Cameron and D. A. Storvick, *An L_2 analytic Fourier-Feynman transform*, Michigan Math. J. **23** (1976), 1-30. MR **53**:8371
- [2] ———, *Feynman integral of variations of functions, in Gaussian random fields*, Ser. Prob. Statist. **1** (1991), 144-157. MR **93b**:28035
- [3] K. S. Chang, B. S. Kim, and I. Yoo, *Fourier-Feynman transform, convolution and first variation of functionals on abstract Wiener space*, Integral transforms and Special Functions **10** (2000), 179-200. MR **2001m**:28023
- [4] S. J. Chang and D. L. Skoug, *The effect of drift on the Fourier-Feynman transform, the convolution product and the first variation*, Panamerican Math. J. **10** (2000), 25-38.

- [5] ———, *Generalized Fourier-Feynman transforms and a first variation on function space*, to appear in *Integral Transforms and Special Functions*.
- [6] T. Huffman, C. Park and D. Skoug, *Analytic Fourier-Feynman transforms and convolution*, *Trans. Amer. Math. Soc.* **347** (1995), 661-673. MR **95d**:28017
- [7] ———, *Generalized transforms and convolutions*, *Internat. J. Math. and Math. Sci.* **20** (1997), 19-32. MR **97k**:46047
- [8] G. W. Johnson and D. L. Skoug, *An L_p Analytic Fourier-Feynman transform*, *Michigan Math. J.* **26** (1979), 103-127. MR **81a**:46050
- [9] ———, *Scale-invariant measurability in Wiener space*, *Pacific J. Math* **83** (1979), 157-176. MR **81b**:28016
- [10] E. Nelson, *Dynamical theories of Brownian motion (2nd edition)*, Math. Notes, Princeton University Press, Princeton (1967). MR **35**:5001
- [11] C. Park, and D. Skoug, *Integration by parts formulas involving analytic Feynman integrals*, *Panamerican Math. J.* **8** (1998), 1-11. MR **99i**:46031
- [12] H. L. Royden, *Real Analysis* (Third edition), Macmillan (1988). MR **90g**:00004
- [13] J. Yeh, *Stochastic Processes and the Wiener Integral*, Marcel Dekker, Inc., New York (1973). MR **57**:14166

DEPARTMENT OF MATHEMATICS, DANKOOK UNIVERSITY, CHEONAN 330-714, KOREA
E-mail address: sejchang@dankook.ac.kr

DEPARTMENT OF MATHEMATICS, DANKOOK UNIVERSITY, CHEONAN 330-714, KOREA
E-mail address: jgchoi@dankook.ac.kr

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NEBRASKA, LINCOLN,
NEBRASKA, 68588-0323
E-mail address: dskoug@math.unl.edu

THERMODYNAMIC FORMALISM FOR COUNTABLE TO ONE MARKOV SYSTEMS

MICHIKO YURI

ABSTRACT. For countable to one transitive Markov systems we establish thermodynamic formalism for non-Hölder potentials in nonhyperbolic situations. We present a new method for the construction of conformal measures that satisfy the weak Gibbs property for potentials of weak bounded variation and show the existence of equilibrium states equivalent to the weak Gibbs measures. We see that certain periodic orbits cause a phase transition, non-Gibbsianness and force the decay of correlations to be slow. We apply our results to higher-dimensional maps with indifferent periodic points.

§0. INTRODUCTION

Thermodynamic formalism for hyperbolic systems was satisfactorily established with Bowen's program ([2]). The existence of generating finite Markov partitions and analysis of Ruelle-Perron-Frobenius operators associated to Hölder potentials allow one to show the existence of unique equilibrium states that satisfy the Gibbs property (in the sense of Bowen) and exponential decay of correlations. Also, the pressure functions are analytic and there is no possibility of *phase transition* (non-uniqueness of equilibrium states). Furthermore, the analyticity problem is strongly related to multifractal problems and the zero of Bowen's equation determines the Hausdorff dimension of limit sets arising from certain iterated functional systems ([4], [8], [12], [22]). On the other hand, phase transition, failure of the Gibbs property and slow decay of correlations can be observed for many complex systems which exhibit common phenomena in transition to turbulence (the so-called *Intermittency*). In this paper we shall construct mathematical models which exhibit such phenomena and for this purpose we shall establish thermodynamic formalism for non-Hölder potentials in nonhyperbolic situations in the following sense: generating Markov partitions are countable partitions and dynamical instability is *subexponential* (subexponential decay of cylinder sizes). More specifically, for countable to one transitive Markov systems we shall construct conformal measures ν that are *weak Gibbs* measures for potentials ϕ of *weak bounded variation* (WBV) (see definitions in §1) and show the existence of equilibrium states μ for ϕ equivalent to the weak Gibbs measures ν . The conformal measures ν associated to ϕ play important roles as reference measures from the physical point of view, and absolute continuity of equilibrium states allows one to describe statistical properties of observable phenomena in the physical sense. In order to clarify typical reasons for

Received by the editors March 28, 2002 and, in revised form, September 10, 2002.

2000 *Mathematics Subject Classification.* Primary 28D99, 28D20, 58C40, 58E30, 37A40, 37A30, 37C30, 37D35, 37F10, 37A45.

phase transition, non-Gibbsianness and slow decay of correlations, we introduce in §4 a notion of *indifferent periodic point* associated to potentials ϕ of WBV. Those periodic points cause failure of summable variation for potentials ϕ and failure of bounded distortion of local Jacobians with respect to the weak Gibbs measure ν for ϕ (Proposition 4). Then a construction of weak Gibbs measures ν for ϕ admitting indifferent periodic points implies “subexponential instability” in terms of cylinder measures (Proposition 3). Furthermore, a construction of equilibrium states μ for such ϕ equivalent to ν (Theorem 6, Lemma 15) allows us to show both phase transition (Corollary 2, Theorem 8) and non-Gibbsianness of equilibrium states (Theorem 5). In particular, our results are applicable to the following piecewise C^1 -smooth countable to one Markov maps T defined on bounded regions $X \subset \mathbb{R}^d$ with indifferent periodic points ($T^q x_0 = x_0, |\det DT^q(x_0)| = 1$) for which the potentials $-\log |\det DT|$ satisfy neither summable variation nor bounded distortion, so that previous results cannot be applicable.

Example A. (Inhomogeneous Diophantine approximations [13], [15], [16], [17], [19], [20], [21]). Let $X = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq 1, -y \leq x < -y + 1\}$ and define $T : X \rightarrow X$ by

$$T(x, y) = \left(\frac{1}{x} - \left\lfloor \frac{1-y}{x} \right\rfloor + \left\lfloor -\frac{y}{x} \right\rfloor, -\left\lfloor -\frac{y}{x} \right\rfloor - \frac{y}{x} \right),$$

where $[x] = \max\{n \in \mathbb{Z} \mid n \leq x\}$ ($x \in \mathbb{N}$) and $[x] = \max\{n \in \mathbb{Z} \mid n < x\}$ ($x \in \mathbb{Z} \setminus \mathbb{N}$). This map admits indifferent periodic points $(1, 0)$ and $(-1, 1)$ with period 2, i.e., $|\det DT^2(1, 0)| = |\det DT^2(-1, 1)| = 1$ and is related to a Diophantine approximation problem of inhomogeneous linear class. \square

Example B (A complex continued fraction [5], [12], [15], [22]). We can define a complex continued fraction transformation $T : X \rightarrow X$ on the diamond-shaped region $X = \{z = x_1\alpha + x_2\bar{\alpha} : -1/2 \leq x_1, x_2 \leq 1/2\}$, where $\alpha = 1 + i$, by $T(z) = 1/z - [1/z]_1$. Here $[z]_1$ denotes $[x_1 + 1/2]\alpha + [x_2 + 1/2]\bar{\alpha}$, where z is written in the form $z = x_1\alpha + x_2\bar{\alpha}$, $[x] = \max\{n \in \mathbb{Z} \mid n \leq x\}$ ($x \in \mathbb{N}$) and $[x] = \max\{n \in \mathbb{Z} \mid n < x\}$ ($x \in \mathbb{Z} - \mathbb{N}$). This transformation has an indifferent periodic orbit $\{1, -1\}$ of period 2 and two indifferent fixed points at i and $-i$. \square

We recall previous works related to thermodynamic formalism for countable to one Markov systems. For countable Markov shifts, O. Sarig proved the existence of conformal measures and equilibrium states associated to *locally* Hölder potentials defined in [10] and D. Fiebig, U. Fiebig and the author proved the existence of equilibrium states for potentials satisfying bounded distortion ($\sup_{n \geq 1} C_n < \infty$ in the definition of WBV) in [7]. Our main Theorems 4-8 do not satisfy these assumptions and Examples A, B show that they cannot be treated by methods in [10] and in [7]. Furthermore, for higher-dimensional systems that are not symbolic systems we may have crucial difficulties in verifying the positive recurrence condition imposed on potentials in both [10] and [7]. The infinite iterated functional systems that Mauldin and Urbanski studied in [8] correspond to the local inverses of piecewise conformal countable Bernoulli systems in our sense, and the method used in [8] severely relies on the Bernoulli property which fails to hold for Example B. Moreover, a Hölder-type condition, imposed on potentials for the existence of conformal measures and for establishing a variational principle, is not satisfied by the important potentials $-\log |\det DT|$ for both Examples A and B.

In order to prove our theorems, we first give in §2 an appropriate definition of topological pressure for countable to one transitive Markov systems with *finite range structure*. Our definition coincides with the standard one by using periodic points under certain conditions (Lemma 6) which can be easily verified for higher-dimensional examples in §8. We also associate the topological pressure to the spectral radius of the Perron-Frobenius-Ruelle operator (Theorem 3). The essential issue for constructing both weak Gibbs measures ν for ϕ and equilibrium states μ for ϕ equivalent to ν is to derive Schweiger's *jump transformations* T^* over full cylinders (see the definition in §1) with respect to which a *local exponential instability* (05) and a *local bounded distortion* (06) for potentials ϕ are satisfied. Then showing the existence of a zero of a generalized Bowen's equation (GBE) for derived potentials ϕ^* associated to T^* (Lemma 7) allows one to show the existence of conformal measures that are weak Gibbs measures for ϕ (Theorem 4). Under a mild condition which cannot be covered by previous works, we show the existence of a zero of (GBE) in §3 by using a product formula of zeta functions (Proposition 1), which shows a nice relation between zeta functions for the original systems and zeta functions for the jump transformations. We also construct σ -finite conformal measures via induced maps T_A over a single full cylinder A (Theorem 7) in §6 by using some idea that appeared in a previous work by M. Denker and the author [5] in which no evidence of the existence of weak Gibbs conformal measures was given. We establish the existence of equilibrium states μ for ϕ of WBV equivalent to the weak Gibbs measure ν for ϕ via a jump transformation (Theorem 6) in §5 and via induced maps (Lemma 15) in §6. Then we can immediately see that the appearance of indifferent periodic orbits associated to ϕ implies a phase transition, i.e., failure of the uniqueness of equilibrium states (Corollary 2, Theorem 8). We should remark that our construction via induced maps shows the existence of (countably many) mutually singular equilibrium states. In §8 we apply our results to higher-dimensional piecewise C^1 Markov maps with indifferent periodic points. All proofs of results in §§2-3 are postponed to the Appendix.

§1. PRELIMINARIES

Let (X, d) be a compact metric space and let $T : X \rightarrow X$ be a noninvertible map that is not necessarily continuous. Suppose that there exists a countable disjoint partition $Q = \{X_i\}_{i \in I}$ of X such that $\bigcup_{i \in I} \text{int} X_i$ is dense in X and the following properties are satisfied.

- (01) For each $i \in I$ with $\text{int} X_i \neq \emptyset$, $T|_{\text{int} X_i} : \text{int} X_i \rightarrow T(\text{int} X_i)$ is a homeomorphism and $(T|_{\text{int} X_i})^{-1}$ extends to a homeomorphism v_i on $cl(T(\text{int} X_i))$.
- (02) $T(\bigcup_{\text{int} X_i = \emptyset} X_i) \subset \bigcup_{\text{int} X_i = \emptyset} X_i$.
- (03) $\{X_i\}_{i \in I}$ generates \mathcal{F} , the σ -algebra of Borel subsets of X .

We say that the triple $(T, X, Q = \{X_i\}_{i \in I})$ is a piecewise C^0 -invertible system. By (01), $T|_{\text{int} X_i}$ extends to a homeomorphism $(v_i)^{-1}$ on $cl(\text{int} X_i)$ for $i \in I$ with $\text{int} X_i \neq \emptyset$. For notational convenience we denote $(v_i)^{-1} = T|_{cl(\text{int} X_i)}$. Let $\underline{i} = (i_1 \dots i_n) \in I^n$ satisfy $\text{int}(X_{i_1} \cap T^{-1} X_{i_2} \cap \dots \cap T^{-(n-1)} X_{i_n}) \neq \emptyset$. Then we define $X_{\underline{i}} := X_{i_1} \cap T^{-1} X_{i_2} \cap \dots \cap T^{-(n-1)} X_{i_n}$, which is called a cylinder of rank n and write $|\underline{i}| = n$. By (01), $T^n|_{\text{int} X_{i_1 \dots i_n}} : \text{int} X_{i_1 \dots i_n} \rightarrow T^n(\text{int}(X_{i_1 \dots i_n}))$ is a homeomorphism and $(T^n|_{\text{int} X_{i_1 \dots i_n}})^{-1}$ extends to a homeomorphism $v_{i_1} \circ v_{i_2} \circ \dots \circ v_{i_n} = v_{i_1 \dots i_n} : cl(T^n(\text{int} X_{\underline{i}})) \rightarrow cl(\text{int} X_{\underline{i}})$ and $(v_{i_1 \dots i_n})^{-1} = T^n|_{cl(\text{int} X_{i_1 \dots i_n})}$. We impose

on (T, X, Q) the next condition, which gives a nice countable states symbolic dynamics similar to sofic shifts (cf. [15], [16], [17], [19]):

(*Finite Range Structure*). $\mathcal{U} = \{\text{int}(T^n X_{i_1 \dots i_n}) : \forall X_{i_1 \dots i_n}, \forall n > 0\}$ consists of finitely many open subsets U_1, \dots, U_N of X . \square

In particular, if (T, X, Q) satisfies the Markov property (i.e., $\text{int}X_i \cap \text{int}TX_j \neq \emptyset$ implies $\text{int}TX_j \supset \text{int}X_i$), then $\mathcal{U} = \{\text{int}(TX_i) : \forall i \in I\}$ and we say that (T, X, Q) is an *FRS Markov system*. If $X_i \in Q$ satisfies $\text{cl}(T(\text{int}X_i)) = X$, then X_i is called a *full cylinder*. If all cylinders are full cylinders so that $\mathcal{U} = \{\text{int}X\}$, then (T, X, Q) is called a *Bernoulli system*. We assume further the next transitive condition:

(*Transitivity*). $\text{int}X = \bigcup_{k=1}^N U_k$ and $\forall l \in \{1, 2, \dots, N\}, \exists 0 < s_l < \infty$ such that for each $k \in \{1, 2, \dots, N\}$, U_k contains an interior of a cylinder $X^{(k,l)}(s_l)$ of rank s_l such that $T^{s_l}(\text{int}X^{(k,l)}(s_l)) = U_l$. \square

The transitivity condition allows one to establish the next fact.

Lemma 1. *There exists $0 < S < \infty$ such that $T^S(\bigcup_{l=1}^N \text{int}X^{(k,l)}(s_l)) = \text{int}X$ and $\forall X_i \in Q, T^{S+1}(\text{int}X_i) = \text{int}X$.*

Proof of Lemma 1. Since each U_k contains $\bigcup_{l=1}^N \text{int}X^{(k,l)}(s_l)$, choosing $S = \prod_{l=1}^N s_l$ is enough to establish the desired fact. \square

Remark (A). If (T, X, Q) is a Markov system, then the transitivity condition implies *aperiodicity* in the following sense: $\exists S > 0$ such that $\forall U_k, U_l \in \mathcal{U}, \forall n > S, \exists \underline{i} = (i_1 \dots i_n)$ with $\text{int}X_{\underline{i}} \neq \emptyset$ satisfying $\text{int}X_{i_1} \subset U_k$ and $T(\text{int}(X_{i_n})) = U_l$.

Definition. We say that ϕ is a potential of *weak bounded variation* (WBV) if there exists a sequence of positive numbers $\{C_n\}$ satisfying $\lim_{n \rightarrow \infty} (1/n) \log C_n = 0$ and $\forall n \geq 1, \forall X_{i_1 \dots i_n} \in \bigvee_{j=0}^{n-1} T^{-j}Q$,

$$\frac{\sup_{x \in X_{i_1 \dots i_n}} \exp(\sum_{j=0}^{n-1} \phi(T^j x))}{\inf_{x \in X_{i_1 \dots i_n}} \exp(\sum_{j=0}^{n-1} \phi(T^j x))} \leq C_n.$$

Define

$$\text{Var}_n(T, \phi) := \sup_{Y \in \bigvee_{j=0}^{n-1} T^{-j}(Q)} \sup_{x, y \in Y} |\phi(x) - \phi(y)|.$$

Remark (B). If $\text{Var}_n(T, \phi) \rightarrow 0$ as $n \rightarrow \infty$, which implies continuity of ϕ in symbolic distance, then ϕ satisfies the WBV property. Hence if (T, X, Q) is a subshift of finite type, then any continuous functions satisfy the WBV property and if (T, X, Q) is a countable Markov shift, then any uniformly continuous functions ϕ with $\text{Var}_1(T, \phi) < \infty$ satisfy the WBV property ([7]).

Let \mathcal{F} be the σ -algebra of Borel sets of the compact space X .

Definition ([17], [18], [20]). A probability measure ν on (X, \mathcal{F}) is called a *weak Gibbs measure* for a function ϕ with a constant P if there exists a sequence $\{K_n\}_{n>0}$ of positive numbers with $\lim_{n \rightarrow \infty} (1/n) \log K_n = 0$ such that ν -a.e. x ,

$$K_n^{-1} \leq \frac{\nu(X_{i_1 \dots i_n}(x))}{\exp(\sum_{j=0}^{n-1} \phi T^j(x) + nP)} \leq K_n,$$

where $X_{i_1 \dots i_n}(x)$ denotes the cylinder containing x .

For a function $\phi : X \rightarrow \mathbb{R}$, we define an operator \mathcal{L}_ϕ by

$$\mathcal{L}_\phi g(x) = \sum_{i \in I} \exp \phi(v_i(x)) g(v_i(x)) 1_{cl(T(int X_i))}(x) \quad (\forall g \in C(X), \forall x \in X).$$

If ϕ satisfies $Var_n(\phi) \rightarrow 0$ ($n \rightarrow \infty$), $\|\mathcal{L}_\phi 1\| := \sup_{x \in X} \mathcal{L}_\phi 1(x) < \infty$ and

(04) $\{v_i\}_{i \in I}$ is an equi-continuous family of partially defined uniformly continuous maps,

then \mathcal{L}_ϕ preserves $C(X)$ (i.e., $\mathcal{L}_\phi : C(X) \rightarrow C(X)$) and is called the *Ruelle-Perron-Frobenius operator*. We remark that (04) is valid if $\sigma(n) = \sup\{\text{diam } Y \mid Y \in \bigvee_{j=0}^{n-1} T^{-j}(Q)\} \rightarrow 0$ as $n \rightarrow \infty$.

We recall the next result, which follows from Theorem 5.1 in [17] and Proposition 2.2 in [18].

Lemma 2 ([17], [18]). *Let (T, X, Q) be a transitive FRS Markov system satisfying $int X \in \mathcal{U}$, and let ϕ be a potential of WBV. Assume that there exist $p > 0$ and a Borel probability measure ν on (X, \mathcal{F}) satisfying $\mathcal{L}_\phi^* \nu = p\nu$, where \mathcal{L}_ϕ^* is the dual of \mathcal{L}_ϕ . Then ν is a weak Gibbs measure for ϕ with $-\log p$.*

Definition. We say that a Borel probability measure ν on X is an *f-conformal measure* if $\frac{d(\nu T)|_{X_i}}{d\nu|_{X_i}} = f|_{X_i}$ ($\forall i \in I$) and $\nu(\bigcup_{i \in I} \partial X_i) = 0$.

In order to show the weak Gibbs property of ν , we use the following formula of the local Jacobians with respect to ν :

$$\frac{d(\nu T)|_{X_i}}{d\nu|_{X_i}} = \exp[\log p - \phi]|_{X_i} \quad (\forall i \in I).$$

Thus for the existence of weak Gibbs measures, it is enough to show the existence of conformal measures (see §3).

Lemma 3 (Theorem 2.2 in [18]). *Let ν be a weak Gibbs measure for ϕ with $-P$. If there exists a T -invariant ergodic probability measure μ equivalent to ν with $H_\mu(Q) < \infty$ and $\phi \in L^1(\mu)$, then $P = h_\mu(T) + \int_X \phi d\mu$.*

In particular, if the constant P is the measure-theoretical pressure, then the existence of a T -invariant ergodic probability measure μ equivalent to the weak Gibbs measure ν for ϕ with $-P$ implies the existence of an equilibrium state for ϕ (see §4). In order to achieve both constructions of conformal measures and equilibrium states, we need to introduce new derived systems which are called *jump transformations* ([13]). Let $B_1 \subset X$ be a union of cylinders of rank 1 of which index i belongs to a subset J of I , and let $D_1 := B_1^c$. Define a function $R : X \rightarrow \mathbb{N} \cup \{\infty\}$ by $R(x) = \inf\{n \geq 0 : T^n x \in B_1\} + 1$ and for each $n > 1$, define inductively

$$B_n := \{x \in X \mid R(x) = n\}, \quad D_n := \{x \in X \mid R(x) > n\} = \bigcap_{m=0}^{n-1} T^{-m} B_1^c.$$

Now we define Schweiger's jump transformation ([13]) $T^* : \bigcup_{n=1}^\infty B_n \rightarrow X$ by $T^* x = T^{R(x)} x$. We denote $X^* := X \setminus (\bigcup_{m=0}^\infty T^{*-m}(\bigcap_{n \geq 0} \{R(x) > n\}))$ and

$$I^* := \bigcup_{n \geq 1} \{(i_1 \dots i_n) \in I^n : X_{i_1 \dots i_n} \subseteq B_n\}.$$

Then it is easy to see that $(T^*, X^*, Q^* = \{X_i\}_{i \in I^*})$ is an FRS Markov system. For a given $\phi : X \rightarrow \mathbb{R}$, we define $\phi^* : \bigcup_{n=1}^\infty B_n \rightarrow \mathbb{R}$ by $\phi^*(x) = \sum_{h=0}^{R(x)-1} \phi T^h(x)$.

Definition. We say that an FRS Markov system satisfies *local exponential instability with respect to* B_1 if (05) : $\exists 0 < \gamma^* < 1, \exists 0 < \Gamma^* < \infty$ such that $\forall n \geq 1$,

$$\sigma_{X^*, T^*}(n) = \sup\{\text{diam } Y \mid Y \in \bigvee_{j=0}^{n-1} T^{*-j}(Q^*)\} \leq \Gamma^* \gamma^{*n}.$$

Definition. We say that a potential $\phi : X \rightarrow \mathbb{R}$ satisfies *local bounded distortion with respect to* B_1 if there exists $\theta > 0$ such that (06) : $\forall \underline{i} = (i_1 \dots i_{|\underline{i}|}) \in I^*, \exists 0 < L_\phi(\underline{i}) < \infty$ satisfying

$$|\phi(v_{\underline{i}}(x)) - \phi(v_{\underline{i}}(y))| \leq L_\phi(\underline{i}) d(x, y)^\theta \quad (\forall x, y \in T^{|\underline{i}|} X_{\underline{i}})$$

and

$$\sup_{\underline{i} \in I^*} \sum_{j=0}^{|\underline{i}|-1} L_\phi(i_{j+1} \dots i_{|\underline{i}|}) < \infty.$$

Under the conditions (05-06), we can easily verify that $\{\phi^* v_{\underline{i}} : \underline{i} \in I^*\}$ is an equi-Hölder continuous family (cf. [19], [20]) and $\sum_{n=1}^\infty \text{Var}_n(T^*, \phi^*) < \infty$. Both conditions (05-06) can be easily verified for all higher-dimensional examples in §8.

In the rest of this section we shall state relations between jump transformations associated to B_1 and induced maps over B_1 . Let $R_{B_1} : B_1 \rightarrow \mathbb{R} \cup \{\infty\}$ be the first return function defined by $R_{B_1}(x) = \inf\{n \geq 1 : T^n x \in B_1\}$. Then we define the induced map T_{B_1} over $\{x \in B_1 : R_{B_1}(x) < \infty\}$ by $T_{B_1} x = T^{R_{B_1}(x)} x$ and the induced potential $\phi_{B_1} : \{x \in B_1 : R_{B_1}(x) < \infty\} \rightarrow \mathbb{R}$ by $\phi_{B_1}(x) = \sum_{h=0}^{R_{B_1}(x)-1} \phi T^h(x)$. Then we can immediately see the following facts.

Lemma 4. (4-1) $R_{B_1} = R \circ T|_{B_1}$,

$$(4-2) \quad \phi_{B_1} - sR_{B_1} = (\phi^* - sR) \circ T|_{B_1} + (\phi - \phi \circ T_{B_1}),$$

$$(4-3) \quad \sum_{m=0}^{n-1} (\phi_{B_1} - sR_{B_1}) \circ T_{B_1}^m = \sum_{m=0}^{n-1} (\phi^* - sR) \circ T^{*m} \circ T|_{B_1^*} + (\phi - \phi \circ T_{B_1}^n),$$

where $B_1^* := B_1 \setminus (\bigcup_{m=0}^\infty T_{B_1}^{-m}(\{R_{B_1}(x) = \infty\}))$.

Lemma 5 (Lemma 4.1 in [18]). Suppose that B_1 consists of full cylinders. Then for any T -invariant probability measure m with $m(B_1) > 0$, $m_{B_1} := \frac{m|_{B_1}}{m(B_1)}$ is a T_{B_1} -invariant probability measure and $m^* := m_{B_1} T|_{B_1}^{-1}$ is a T^* -invariant probability measure. m can be written in terms of m^* by Schweiger's formula (see (3) in §5) and in terms of m_{B_1} by Kac's formula (see Lemma 16).

§2. TOPOLOGICAL PRESSURE FOR POTENTIALS OF WEAK BOUNDED VARIATION

Let (T, X, Q) be a transitive FRS Markov system and let $\phi : X \rightarrow \mathbb{R}$ be a potential of WBV. For each $n > 0$ and for each $U \in \mathcal{U}$ we define the following partition functions :

$$Z_n(U, \phi) := \sum_{\underline{i}: |\underline{i}|=n, \text{int}(TX_{i_n})=U \supset \text{int} X_{i_1}} \sum_{v_{\underline{i}} x = x \in \text{cl}(\text{int} X_{\underline{i}})} \exp\left[\sum_{h=0}^{n-1} \phi T^h(x)\right],$$

$$\overline{Z}_n(U, \phi) = \sum_{\underline{i}: |\underline{i}|=n, \text{int}(TX_{i_n})=U \supset \text{int} X_{i_1}} \sup_{x \in X_{\underline{i}}} \exp\left[\sum_{h=0}^{n-1} \phi T^h(x)\right]$$

and

$$\underline{Z}_n(U, \phi) = \sum_{\underline{i}: |\underline{i}|=n, \text{int}(TX_{i_n})=U \supset \text{int}X_{i_1}} \inf_{x \in X_{\underline{i}}} \exp\left[\sum_{h=0}^{n-1} \phi T^h(x)\right].$$

We further define

$$Z_n(\phi) := \sum_{\underline{i}: |\underline{i}|=n, \text{int}(TX_{i_n}) \supset \text{int}X_{i_1}} \sum_{v_{\underline{i}} x = x \in \text{cl}(\text{int}X_{\underline{i}})} \exp\left[\sum_{h=0}^{n-1} \phi T^h(x)\right].$$

We shall define the topological pressure as the asymptotic growth rates of these partition functions.

Theorem 1 (Topological pressure for potentials of WBV). *Let (T, X, Q) be a transitive FRS Markov system and let ϕ be a potential of WBV. For each $U \in \mathcal{U}$, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \overline{Z}_n(U, \phi)$, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \underline{Z}_n(U, \phi)$, $\lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(U, \phi)$ exist and do not depend on U . Furthermore, the limits coincide with $\lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\phi)$.*

We call the limit $P_{\text{top}}(T, \phi) := \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\phi)$ the *topological pressure* for ϕ . The next fact can be verified easily.

Lemma 6. *Under the next condition, $Z_n(\phi)$ coincides with the usual partition function defined by : $\sum_{T^n x = x, x \in X} \exp[\sum_{h=0}^{n-1} \phi T^h(x)]$.*

- (1) *For $x_0 \in X_{i_1 \dots i_n}$ with $T^n x_0 = x_0$, either $x_0 \in \text{int}X_{i_1 \dots i_n}$ or $x_0 \notin \text{cl}X_{j_1 \dots j_n}$ for $(j_1 \dots j_n) \neq (i_1 \dots i_n)$.*

Let \mathcal{V} be the finite disjoint partition generated by \mathcal{U} . We should claim that if a periodic point x_0 with period n is contained in a cylinder $X_{i_1 \dots i_n}$ satisfying $X_{i_1 \dots i_n} \subset \text{int}V$ for some $V \in \mathcal{V}$, then $x_0 \notin \partial X_{i_1 \dots i_n}$. If not, we have a contradiction to $x_0 \in \text{int}V$ because of $x_0 \in T^n(\partial X_{i_1 \dots i_n}) = \partial(T^n X_{i_1 \dots i_n})$. By using this fact, we will see that all higher-dimensional examples in §8 satisfy (1). The Artin-Mazur-Ruelle zeta function $\zeta_{T, \phi}(z)$ is defined by $\zeta_{T, \phi}(z) = \exp[\sum_{n=1}^{\infty} \frac{z^n}{n} Z_n(\phi)]$. Then the radius of convergence of $\zeta_{T, \phi}(z)$ is given by $\rho_{\phi} = \exp[\limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\phi)]^{-1}$. We define

$$\mathcal{W}(T) := \{\phi : X \rightarrow \mathbb{R} \mid \phi \text{ satisfies WBV and } P_{\text{top}}(T, \phi) < \infty\}$$

and

$$\mathcal{W}_B(T) := \{\phi \in \mathcal{W}(T) \mid \text{Var}_n \phi \rightarrow 0 \ (n \rightarrow \infty), \|\phi\| := \sup_{x \in X} |\phi(x)| < \infty\}.$$

We can easily see that the pressure function $P_{\text{top}}(T, \cdot) : \mathcal{W}(T) \rightarrow \mathbb{R}$ satisfies continuity, convexity and $\forall \phi_1, \phi_2 \in \mathcal{W}(T)$, $P_{\text{top}}(T, \phi_1 + \phi_2) \leq P_{\text{top}}(T, \phi_1) + P_{\text{top}}(T, \phi_2)$. Furthermore, by applying Theorem 2.4 in [7] we have the following fact.

Theorem 2. $\mathcal{W}_B(T)$ is a Banach space and $P_{\text{top}}(T, \cdot) : \mathcal{W}_B(T) \rightarrow \mathbb{R}$ is a Lipschitz continuous convex function.

Definition. If an FRS Markov system (T, X, Q) satisfies that $\forall U \in \mathcal{U}, \exists X_i \in Q$ such that $X_i \subset U$ and $T(\text{int}X_i) = \text{int}X$, then (T, X, Q) is called a *strongly transitive* FRS Markov system.

Theorem 3 (Topological pressure and the spectral radius). *Let (T, X, Q) be a strongly transitive FRS Markov system satisfying $\mathcal{U} \cap \mathcal{V} \neq \emptyset$. Let ϕ be a potential of weak bounded variation. Then $\forall U \in \mathcal{U} \cap \mathcal{V}$ and $\forall x \in U$, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{L}_{\phi}^n 1_U(x) = P_{\text{top}}(T, \phi)$. Furthermore, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_{\phi}^n 1\| = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_{\phi}^n\| = P_{\text{top}}(T, \phi)$.*

We can easily verify all conditions in Theorem 3 for examples in §8.

§3. THE CONSTRUCTION OF WEAK GIBBS CONFORMAL MEASURES

Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. Suppose that there exists a union of full cylinders $B_1(\subset X)$ with respect to which (T, X, Q) satisfies local exponential instability and ϕ satisfies local bounded distortion. For the derived potential $\phi^*(x) = \sum_{h=0}^{R(x)-1} \phi T^h(x)$ we define

$$Z_n(\phi^*) := \sum_{(i_1 \dots i_n) \in I^{*n} : \text{int}(T^* X_{i_n}) \supset \text{int} X_{i_1}} \sum_{v_{i_1} \dots v_{i_n} \ x = x \in \text{cl}(\text{int} X_{i_1 \dots i_n})} \exp \left[\sum_{m=0}^{n-1} \phi T^{*m}(x) \right].$$

Then by Theorem 1, summable variations of ϕ^* allow one to show that

$$\exists \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\phi^*) := P_{\text{top}}(T^*, \phi^*) \in (-\infty, \infty].$$

Theorem 4 (A construction of conformal measures via jump transformations). *Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. Suppose that there exists a union of full cylinders $B_1(\subset X)$ with respect to which (T, X, Q) satisfies local exponential instability and ϕ satisfies local bounded distortion. Assume further that $\|\mathcal{L}_{\phi^* - R \min\{0, P_{\text{top}}(T^*, \phi^*)\}} 1\| < \infty$. Then there exists a Borel probability measure ν on X supported on X^* satisfying*

$$\frac{d\nu T}{d\nu}|_{X_i} = \exp[P_{\text{top}}(T, \phi) - \phi](\forall i \in I)$$

and $\nu(\bigcup_{i \in I} \partial X_i) = 0$.

As we have announced in §0, for constructing a weak Gibbs measure for ϕ of WBV, we shall consider the following generalized Bowen's equation:

$$(\text{GBE}) : P_{\text{top}}(T^*, \phi^* - sR) = 0.$$

The existence of a zero of the equation (GBE) follows from the standard argument in the case when $0 \leq P_{\text{top}}(T^*, \phi^*) < \infty$ because of continuity of the function $s \rightarrow P_{\text{top}}(T^*, \phi^* - sR)$ on $\text{int}\{s \in \mathbb{R} \mid P_{\text{top}}(T^*, \phi^* - sR) \in \mathbb{R}\}$ (see Lemma 14). We should notice that the uniqueness of the zero of (GBE) follows from the "strictly" decreasing property of the function, $s \rightarrow P_{\text{top}}(T^*, \phi^* - sR)$ in the standard situation. Here we have no evidence of it although the function is decreasing. If $P_{\text{top}}(T^*, \phi^*) < 0$, then under the assumption $P_{\text{top}}(T^*, \phi^* - RP_{\text{top}}(T^*, \phi^*)) < \infty$ we see that

$$P_{\text{top}}(T^*, \phi^* - RP_{\text{top}}(T^*, \phi^*)) \geq P_{\text{top}}(T^*, \phi^* - P_{\text{top}}(T^*, \phi^*)) = 0$$

and so we can reduce to the previous case. If $P_{\text{top}}(T^*, \phi^*) = \infty$, then we cannot use the standard argument. Now we come to state the next key lemma, which allows one to establish Theorem 4.

Lemma 7 (The existence of a zero of (GBE)). (i) *If $0 \leq P_{\text{top}}(T^*, \phi^*) < \infty$, then $P_{\text{top}}(T, \phi) \geq 0$ and $\exists s_0 \geq 0$ satisfying $P_{\text{top}}(T^*, \phi^* - s_0 R) = 0$.*

(ii) *If $P_{\text{top}}(T^*, \phi^*) < 0$ and $P_{\text{top}}(T^*, \phi^* - RP_{\text{top}}(T^*, \phi^*)) < \infty$, then $P_{\text{top}}(T, \phi - P_{\text{top}}(T^*, \phi^*)) \geq 0$ and $\exists s_0 \geq 0$ satisfying $P_{\text{top}}(T^*, (\phi - P_{\text{top}}(T^*, \phi^*))^* - s_0 R) = 0$.*

(iii) *If $\sup\{s \in \mathbb{R} : P_{\text{top}}(T^*, \phi^* - sR) = \infty\} = \min\{0, P_{\text{top}}(T^*, \phi^*)\}$, then $\exists s_0 \geq \min\{0, P_{\text{top}}(T^*, \phi^*)\}$ such that $P_{\text{top}}(T^*, \phi^* - s_0 R) = 0$.*

We recall the formal power series $\zeta_{T,\phi}(z) = \exp[\sum_{n=1}^{\infty} \frac{z^n}{n} Z_n(\phi)]$, which is called the Artin-Mazur-Ruelle zeta function. The next product formula of zeta functions plays an important role in proving Lemma 7.

Proposition 1 (cf. [17]). *We can write*

$$(2) \qquad \zeta_{T,\phi}(\exp(-s)) = \zeta_{T^*,\phi^*-sR}(1) \times \zeta_{T|_{\cap_{n>0} D_n},\phi}(\exp(-s)).$$

Corollary 1. *If $s > P_{\text{top}}(T, \phi)$, then $P_{\text{top}}(T^*, \phi^* - sR) \leq 0$.*

By Theorem 3, the assumption $\|\mathcal{L}_{\phi^*-R\min\{0,P_{\text{top}}(T^*,\phi^*)\}}1\| < \infty$ implies either $0 \leq P_{\text{top}}(T^*, \phi^*) < \infty$ or $P_{\text{top}}(T^*, \phi^* - RP_{\text{top}}(T^*, \phi^*)) < \infty$ is satisfied. Hence it follows from Lemma 7 that $\exists s_0 \geq \min\{0, P_{\text{top}}(T^*, \phi^*)\}$ satisfying $P_{\text{top}}(T^*, \phi^* - s_0R) = 0$ and $\|\mathcal{L}_{\phi^*-s_0R}1\| < \infty$. Since $Q^* = \{X_i\}_{i \in I^*}$ consists of full cylinders and summability of variations $\sum_{n=1}^{\infty} Var_n(T^*, \phi^* - s_0R) < \infty$ is valid, we can apply P. Walter’s argument in [14] to show the existence of an $\exp[s_0R - \phi^*]$ -conformal measure with respect to T^* .

Lemma 8. *There exists a Borel probability measure ν on X satisfying $\mathcal{L}_{\phi^*-s_0R}^*\nu = \nu$ and $\nu(\text{int}X^*) = 1$.*

In §9 we shall show the existence of an $\exp[s_0 - \phi]$ -conformal measure for the zero s_0 of (GBE) by using the conformal measure ν on X^* and show $s_0 = P_{\text{top}}(T, \phi)$, which implies uniqueness of the zero of (GBE). At the end of this section, we shall consider the case when $\|\mathcal{L}_{\phi^*-R\min\{0,P_{\text{top}}(T^*,\phi^*)\}}1\| = \infty$. By Theorem 3, if $P_{\text{top}}(T^*, \phi^* - s_0R) = 0$, then there exists sufficiently large n such that $\|\mathcal{L}_{(\phi-s_0)^*_n}^n1\| = \|\mathcal{L}_{(\phi-s_0)^*_n}1\| < \infty$, where

$$(\phi - s_0)^*_n := \sum_{m=0}^{n-1} (\phi - s_0)^* T^{*m} = \sum_{m=0}^{n-1} (\phi^* - s_0R) T^{*m}.$$

We shall introduce a new stopping time (depending on $n \geq 1$) defined on X^* by

$$R_n(x) := \inf\{k \geq n \mid X_{i_1 \dots i_k}(x) \in \bigvee_{m=0}^{n-1} T^{*-m}Q^*\} = \sum_{m=0}^{n-1} R(T^{*m}(x)).$$

Then a new jump transformation S^* defined by $S^*(x) := T^{R_n(x)}(x)$ is equal to T^{*n} and the next facts can be verified easily.

Lemma 9. (9-1) $R_n(T(x)) = R_n(x) - 1$.

$$(9-2) \quad \{x \in X^* \mid R_n(x) = k\} = \{x \in X^* \mid \sum_{m=1}^n |i_m| = k \text{ where } x \in X_{i_1 \dots i_n} \in \bigvee_{m=0}^{n-1} T^{*-m}Q^*\}.$$

$$(9-3) \quad (\phi - s_0)^*_n(x) = \phi^*_n(x) - s_0R_n(x) = \sum_{h=0}^{R_n(x)-1} (\phi - s_0)T^h(x).$$

Now we shall consider a two-parameter family of functions $\{(\phi - s)^*_n \mid (s, n) \in \mathbb{R} \times \mathbb{N}\}$ and the equations $P_{\text{top}}(T^{*n}, (\phi - s)^*_n) = 0$. Applying Theorem 4 gives the next result.

Proposition 2. *Suppose that all conditions in Theorem 3 are satisfied. If there exist $s_0 \in \mathbb{R}$ and $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0, P_{\text{top}}(T^{*n}, (\phi - s_0)^*_n) = 0$, then there exists a Borel probability measure ν on X supported on X^* satisfying $\frac{d\nu T}{d\nu}|_{X_i} = \exp[s_0 - \phi](\forall i \in I)$ and $\nu(\bigcup_{i \in I} \partial X_i) = 0$. Furthermore, $s_0 = P_{\text{top}}(T, \phi)$.*

§4. INDIFFERENT PERIODIC POINTS AND NON-GIBBSIANNES

Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. The next lemma follows from the definition of $P_{\text{top}}(T, \phi)$ directly.

Lemma 10. $P_{\text{top}}(T, \phi) \geq \frac{1}{q} \sum_{h=0}^{q-1} \phi T^h(x_0) (\forall x_0 \in X, T^q x_0 = x_0)$.

Definition. x_0 is called a *generalized indifferent periodic point* with period q with respect to ϕ if $P_{\text{top}}(T, \phi) = \frac{1}{q} \sum_{h=0}^{q-1} \phi T^h(x_0)$. If x_0 is not indifferent, then we call x_0 a *generalized repelling periodic point*.

If a potential ϕ of WBV admits a generalized indifferent periodic point, then we can observe interesting statistical phenomena. More specifically, if there exists an $\exp[P_{\text{top}}(T, \phi) - \phi]$ -conformal measure ν , then the above definitions can be described in terms of the local Jacobians with respect to ν , that is,

$$\frac{d(\nu T^q)}{d\nu}|_{X_{i_1 \dots i_q}(x_0)}(x_0) = \exp[qP_{\text{top}}(T, \phi) - \sum_{h=0}^{q-1} \phi T^h(x_0)] = 1.$$

Then we have the following facts.

Proposition 3. *Let x_0 be a generalized indifferent periodic point with period q with respect to $\phi \in \mathcal{W}(T)$. Let ν be an $\exp[P_{\text{top}}(T, \phi) - \phi]$ -conformal measure. Then*

- (i) $\forall s \geq 1, P_{\text{top}}(T, s\phi) = sP_{\text{top}}(T, \phi)$ and $\forall s < 1, P_{\text{top}}(T, s\phi) \geq sP_{\text{top}}(T, \phi)$.
- (ii) $\nu(X_{i_1 \dots i_n}(x_0))$ decays subexponentially fast.

Proof. By Lemma 10, we have $P_{\text{top}}(T, s\phi) \geq s \frac{1}{q} \sum_{i=0}^{q-1} \phi T^i(x_0)$. In particular, if x_0 is a generalized indifferent periodic point for ϕ , then $P_{\text{top}}(T, s\phi) \geq sP_{\text{top}}(T, \phi)$. We recall that by Lemma 2 the conformal measure ν is a weak Gibbs measure for ϕ of WBV. Then we have for $s \geq 1$,

$$\frac{1}{n} \log Z_n(s\phi) \leq sP_{\text{top}}(T, \phi) + \frac{1}{n} \log C_n K_n^s + \frac{1}{n} \log \sum_{i_1 \dots i_n} \nu(X_{i_1 \dots i_n})^s,$$

where both C_n and K_n satisfy $\lim_{n \rightarrow \infty} \frac{1}{n} \log C_n = 0$ and $\lim_{n \rightarrow \infty} \frac{1}{n} \log K_n = 0$. Since $\lim_{n \rightarrow \infty} \frac{1}{n} \log C_n K_n^s = 0$, we have $P_{\text{top}}(T, s\phi) \leq sP_{\text{top}}(T, \phi)$ for $s \geq 1$. (ii) follows from Proposition 6.1 in [21]. □

Let us recall that ν was obtained by constructing a jump transformation in Theorem 4. Then we can associate the generalized indifferent periodic points to the marginal sets $\bigcap_{n \geq 0} D_n$.

Proposition 4. *Let x_0 be a generalized indifferent periodic point with period q with respect to $\phi \in \mathcal{W}(T)$.*

- (i) (*Failure of bounded distortion*)

$$C_{nq}(x_0) := \sup_{x, y \in X_{i_1 \dots i_{nq}}(x_0)} \frac{\exp[\sum_{h=0}^{nq-1} \phi T^h(x)]}{\exp[\sum_{h=0}^{nq-1} \phi T^h(y)]} \rightarrow \infty$$

monotonically as $n \rightarrow \infty$.

- (ii) $x_0 \in \bigcap_{n \geq 0} D_n$.

Proof. Since $C_n(x_0)$ is the distortion of $\frac{d(\nu T^n)}{d\nu}$ over cylinders $X_{i_1 \dots i_n}(x_0)$, (i) follows from Lemma 6.1 in [21]. Suppose $x_0 \notin \bigcap_{n \geq 0} D_n$. Then by Sublemma A (see §9) we have $x_0 \in X^*$. Since $\sum_{n=1}^{\infty} \text{Var}_n(T^*, \phi^*) < \infty$ implies that $C_{nq}(x_0)$ cannot increase monotonically, we have a contradiction to (i). We complete the proof. \square

Remark (C). We claim that $\bigcap_{n \geq 0} D_n$ can contain repelling periodic points.

If we have a T -invariant probability measure μ equivalent to ν via Kac's formula (Lemma 16) or Schweiger's formula (3) in §5, then the invariant densities $d\mu/d\nu$ are typically unbounded at indifferent periodic points with respect to ν (Lemma 6.2 in [21]) so that we can see interesting phenomena from a statistical point of view ([19], [21]). For example, under the existence of a generalized indifferent periodic point x_0 with respect to ϕ , the rate of decay of correlation may be slower than $\nu(X_{i_1 \dots i_n}(x_0))$, which decays subexponentially fast by (ii) in Proposition 3. We referee [21] for further details. On the other hand, the Dirac measure m supported on the generalized indifferent periodic orbit with respect to ϕ satisfies $P_{\text{top}}(T, \phi) = h_m(T) + \int_X \phi dm$. Hence if we can establish a variational principle for the topological pressure and can construct a T -invariant measure μ equivalent to the weak Gibbs measure ν for ϕ with $-P_{\text{top}}(T, \phi)$, then by Lemma 3 we see immediately failure of uniqueness of equilibrium states. Furthermore, by the definition of *indifferency* we can show failure of Gibbsianness of equilibrium states for ϕ with generalized indifferent periodic points.

Theorem 5 (Characterization of non-Gibbsianness). *Suppose that a potential ϕ with $P_{\text{top}}(T, \phi) < \infty$ admits a generalized indifferent periodic point x_0 . Then there is no Borel probability measure that is Gibbs for ϕ .*

§5. EQUILIBRIUM STATES FOR POTENTIALS OF WEAK BOUNDED VARIATION

Let (T, X, Q) be a transitive FRS Markov system and let $M_T(X)$ be the set of all T -invariant probability measures on (X, \mathcal{F}) . For $m \in M_T(X)$, I_m denotes the conditional information of Q with respect to $T^{-1}\mathcal{F}$. We denote

$$M_T(X, \phi) := \{m \in M_T(X) \mid I_m + \phi \in L^1(m), \text{ either } h_m(T) < \infty \text{ or}$$

$$\int_X \phi dm > -\infty \text{ is satisfied}\}.$$

Theorem 6. *Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. Suppose that there exists a union of full cylinders $B_1(\subset X)$ with respect to which (T, X, Q) satisfies local exponential instability and ϕ satisfies local bounded distortion. Let ν be the $\exp[P_{\text{top}}(T, \phi) - \phi]$ -conformal measure supported on X^* . Assume further that $\Gamma := \bigcap_{n \geq 0} D_n$ consists of periodic points. If $\int_{X^*} R d\nu < \infty$ and $H_\nu(Q^*) < \infty$, then there exists a T -invariant ergodic probability measure μ equivalent to ν that satisfies the following variational principle:*

$$P_{\text{top}}(T, \phi) = h_\mu(T) + \int_X \phi d\mu = \sup\{h_m(T) + \int_X \phi dm \mid m \in M_T(X, \phi) \text{ is ergodic}\}.$$

If $E_T(X, \phi) := \{m \in M_T(X, \phi) \mid h_m(T) + \int_X \phi dm = P_{\text{top}}(T, \phi)\}$ contains at least two elements, then it implies physically coexistence of different phases, which is so-called "*phase transition*". Phase transition may be related to failure of the Gibbs property of equilibrium states (see Theorem 5).

Corollary 2 (Phase transition). *We assume all conditions in Theorem 6. If Γ consists of generalized indifferent periodic points with respect to ϕ , then the set of equilibrium states for ϕ is the convex hull of μ and the set of invariant Borel probability measures supported on Γ .*

In order to prove Theorem 6, we need a sequence of lemmas. Let $M_{T^*}(X^*)$ denote the set of all Borel probability measures on X^* invariant under T^* . For $\phi^* - sR$ we define

$$M_{T^*}(X^*, \phi^* - sR) := \{m^* \in M_{T^*}(X^*) \mid \text{either } h_{m^*}(T^*) < \infty \text{ or}$$

$$\int_{X^*} (\phi^* - sR) dm^* > -\infty \text{ is satisfied}\}.$$

Let $s_0 = P_{\text{top}}(T, \phi)$. Then P. Walter's method in [14] can apply for T^* and for $\phi^* - s_0R$ so that there exists the unique equilibrium state μ^* equivalent to ν and the following variational principle is valid:

$$\begin{aligned} 0 &= P_{\text{top}}(T^*, \phi^* - s_0R) = h_{\mu^*}(T^*) + \int_{X^*} (\phi^* - s_0R) d\mu^* \\ &= \sup\{m^* \in M_{T^*}(X^*, \phi^* - s_0R) \mid h_{m^*}(T^*) + \int_{X^*} (\phi^* - s_0R) dm^*\}. \end{aligned}$$

Since $\sum_{n=1}^{\infty} \text{Var}_n(T^*, \phi^* - s_0R) < \infty$ implies the bounded distortion property with respect to ν :

$$\sup_{n \geq 1} \sup_{Y \in \mathcal{V}_{j=0}^{n-1}} \sup_{T^{*-j}(Q^*)} \sup_{x, x' \in Y} \frac{\frac{d(\nu T^{*n})}{d\nu}|_Y(x)}{\frac{d(\nu T^{*n})}{d\nu}|_Y(x')} < \infty,$$

we can show ergodicity and Bowen's Gibbs property for μ^* . If $\int_{X^*} R d\mu^* < \infty$, then the next Schweiger's formula ([13]) gives a T -invariant ergodic probability measure μ equivalent to ν that satisfies $\mu(B_1) = (\int_{X^*} R d\mu^*)^{-1} > 0$:

$$(3) \quad \left(\int_{X^*} R d\mu^*\right) \mu(E) = \sum_{n=0}^{\infty} \mu^*(D_n \cap T^{-n}E) = \int_{X^*} \sum_{i=0}^{R(x)-1} 1_E T^i(x) d\mu^*(x)$$

and by Lemma 5 for $f \in L^1(\mu)$,

$$(3)^* \quad \int_X f d\mu = \frac{\int_{X^*} \sum_{i=0}^{R(x)-1} f T^i(x) d\mu^*}{\int_{X^*} R d\mu^*} = \frac{\int_{B_1} \sum_{i=0}^{R_{B_1}(x)-1} f T^i(x) d\mu_{B_1}}{\int_{B_1} R_{B_1} d\mu_{B_1}}$$

(cf. Lemma 4.2 in [18]). Since $\int_{X^*} R d\mu^* < \infty$ gives the equality

$$\int_{X^*} (\phi^* - s_0R) d\mu^* = \int_{X^*} \phi^* d\mu^* - s_0 \int_{X^*} R d\mu^*$$

and $H_{\mu^*}(Q^*) < \infty$ gives $h_{\mu}(T) < \infty$, we can establish the following characterization of the zero s_0 of (GBE).

Lemma 11. *If $\mu^* \in M_{T^*}(X^*)$ is ergodic and satisfies $h_{\mu^*}(T^*) + \int_{X^*} (\phi^* - s_0R) d\mu^* = 0$, $\int_{X^*} R d\mu^* < \infty$ and $H_{\mu^*}(Q^*) < \infty$, then*

$$s_0 = \frac{h_{\mu^*}(T^*) + \int_{X^*} \phi^* d\mu^*}{\int_{X^*} R d\mu^*} = h_{\mu}(T) + \int_X \phi d\mu,$$

where μ is obtained by formula (3).

By Lemma 11 we have a T -invariant ergodic probability measure μ equivalent to ν that satisfies $P_{\text{top}}(T, \phi) = h_\mu(T) + \int_X \phi d\mu$.

Lemma 12 (Lemma 4.4 in [18]). *If a T -invariant probability measure m satisfies $m(B_1) = 0$, then $\Gamma := \bigcap_{n \geq 0} D_n$ is a full measure set with respect to m .*

Proof of Theorem 6. By Lemma 11 for all T -invariant ergodic probability measures m on X with $m(B_1) > 0$ and $m \in M_T(X, \phi)$, we can establish

$$0 = \frac{h_\mu(T) + \int_X \phi d\mu - P_{\text{top}}(T, \phi)}{\mu(B_1)} \geq \frac{h_m(T) + \int_X \phi dm - P_{\text{top}}(T, \phi)}{m(B_1)}.$$

On the other hand, by Lemma 12, any T -invariant ergodic probability measure m on X with $m(B_1) = 0$ satisfies $m(\Gamma) = 1$. In particular, if $\Gamma := \bigcap_{n \geq 0} D_n$ consists of periodic points, then $h_m(T) + \int_X \phi dm = \int_\Gamma \phi dm \leq P_{\text{top}}(T, \phi)$, which completes the proof of Theorem 6. \square

§6. THE CONSTRUCTION OF σ -FINITE CONFORMAL MEASURES VIA INDUCED MAPS

Let (T, X, Q) be a transitive FRS Markov system and $\phi \in \mathcal{W}(T)$. Suppose that (T, X, Q) satisfies local exponential instability and ϕ satisfies local bounded distortion with respect to $B_1 = \bigcup_{j \in J} X_j$ ($J \subset I$). Let $A_j = \text{cl}(\text{int} X_j)$ for $j \in J$ and put $A := \bigcup_{j \in J} A_j$. We define the first return function $R_A : A \rightarrow \mathbb{R} \cup \{\infty\}$ and the induced map T_A over $\{x \in A : R_A(x) < \infty\}$. By the Markov property, there exists a partition of the set $B_k^{(A)} = \{x \in A : R_A(x) = k\}$ for each $k \geq 1$ so that T^k restricted to the interior of each element of the partition is a homeomorphism onto its image. I_A denotes the set of all indices corresponding to such elements of the partition of $\bigcup_{k \geq 1} B_k^{(A)}$. Then $\{v_{\underline{i}} : \underline{i} \in I_A\}$ is a family of extensions of local inverses of T_A . For $s \in \mathbb{R}$ and $x \in \bigcup_{k=1}^\infty B_k^{(A)}$ we define $\phi_A^{(s)}(x) = \sum_{h=0}^{R_A(x)-1} \phi T^h(x) - s R_A(x)$. By Lemma 4 we can easily see the next fact.

Lemma 13. *If each $A_i \subset A$ satisfies $T A_i = X$, then*

$$P_{\text{top}}(T^*, \phi^* - sR) = P_{\text{top}}(T_A, \phi_A^{(s)}).$$

We recall the following result in [6].

Lemma 14 ([6]). *If $\|\mathcal{L}_{\phi_A^{(0)}} 1\| < \infty$, then $s \rightarrow P_{\text{top}}(T_A, \phi_A^{(s)})$ is continuous on $\text{int}\{s \in \mathbb{R} : P_{\text{top}}(T_A, \phi_A^{(s)}) \in \mathbb{R}\}$.*

We suppose that B_1 consists of a single full cylinder X_j and the following conditions are satisfied for $A = \text{cl}(\text{int} X_j)$.

(05)* $\exists 0 < \gamma < 1, 0 < \bar{\gamma} < \infty$ such that

$$\sigma_{A, T_A}(n) = \sup\{\text{diam } Y \mid Y \in \bigvee_{j=0}^{n-1} T_A^{-j}(Q_A)\} \leq \bar{\gamma} \gamma^n$$

and there exists $\theta > 0$ such that

(06)* $\forall \underline{i} = (i_1 \dots i_{|\underline{i}|}) \in I_A$ and all $0 \leq j < |\underline{i}|$, $\exists 0 < L_\phi(i_{j+1} \dots i_{|\underline{i}|}) < \infty$ satisfying

$$|\phi(v_{i_{j+1} \dots i_{|\underline{i}|}}(x)) - \phi(v_{i_{j+1} \dots i_{|\underline{i}|}}(y))| \leq L_\phi(i_{j+1} \dots i_{|\underline{i}|}) d(x, y)^\theta \quad (\forall x, y \in A)$$

and

$$\sup_{\underline{i} \in I_A} \sum_{j=0}^{|\underline{i}|-1} L_\phi(i_{j+1} \dots i_{|\underline{i}|}) < \infty.$$

Since the conditions (05-06)* allow us to establish the WBV property of $\phi_A^{(s)}$, by Theorem 1, $\exists \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\phi_A^{(s)}) := P_{\text{top}}(T_A, \phi_A^{(s)})$, where

$$Z_n(\phi_A^{(s)}) = \sum_{(\underline{i}_1 \dots \underline{i}_n) \in I_A^n} \sum_{v_{\underline{i}_1} \dots v_{\underline{i}_n} x = x} \exp\left[\sum_{m=0}^{n-1} \phi_A^{(s)} T_A^m(x)\right].$$

Furthermore, (05-06)* guarantee equi-Hölder continuity of $\{\phi_A^{(s)} v_{\underline{i}} : \underline{i} \in I_A\}$ and $\sum_{n=1}^{\infty} \text{Var}_n(T_A, \phi_A^{(s)}) < \infty$. Hence if $\|\mathcal{L}_{\phi_A^{(s)}} 1\| < \infty$, then $\mathcal{L}_{\phi_A^{(s)}} : C(A) \rightarrow C(A)$.

Theorem 7 (A construction of σ -finite conformal measures via induced maps). *Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. Suppose that there exists a full cylinder $X_j \in Q$ satisfying (05)* and (06)* for $A := \text{cl}(\text{int} X_j)$. If*

$$\|\mathcal{L}_{\phi_A^{(0)} - R_A \min\{0, P_{\text{top}}(T_A, \phi_A^{(0)})\}} 1\| < \infty,$$

then

- (i) $\exists s_0 \in \mathbb{R}$ with $P_{\text{top}}(T_A, \phi_A^{(s_0)}) = 0$ (a generalized Bowen's equation);
- (ii) there exists a Borel probability measure ν_A on $(A, \mathcal{F} \cap A)$ with $\nu_A(\{x \in A \mid R_A(x) < \infty\}) = 1$ satisfying $\mathcal{L}_{\phi_A^{(s_0)}}^* \nu_A = \nu_A$;
- (iii) there exists a σ -finite measure ν on X satisfying $\mathcal{L}_\phi^* \nu = [\exp s_0] \nu$ and $\nu(\bigcup_{i \in I} \partial X_i) = 0$;
- (iv) in particular, if ν is finite, then $P_{\text{top}}(T, \phi) = s_0$.

Proof of Theorem 7. By Lemmas 13-14 and Proposition 1, we have the existence of $s \in \mathbb{R}$ for which $P_{\text{top}}(T_A, \phi_A^{(s)}) = 0$ and $\mathcal{L}_{\phi_A^{(s)}}^*$ admits an eigenvalue 1. Then we can apply the main theorem in [5] so that (i)-(iv) are obtained. \square

The next result gives a criterion of finiteness of ν .

Proposition 5 (A criterion of finiteness of ν). *Suppose that all assumptions in Theorem 7 are satisfied. Then $\nu(X) = \int_A \exp[s_0 - \phi] d\nu_A + 1$. In particular, if $\inf_{x \in A} \phi(x) > -\infty$, then ν is finite.*

Proof of Proposition 5. Let $I' = \{\underline{i} \mid X_{\underline{i}} \subset D_{|\underline{i}|}\}$: First we note the following formula of ν , which was obtained in [5]:

$$\nu(E) = \sum_{\underline{i} \in I'} \int_A 1_E(v_{\underline{i}} x) \exp\left[\sum_{l=0}^{|\underline{i}|-1} \phi T^l(v_{\underline{i}} x) - s_0 |\underline{i}|\right] d\nu_A(x) + \int_A 1_E(x) d\nu_A(x).$$

Then we see that $\nu(X)$ is equal to

$$\sum_{k=1}^{\infty} \sum_{\underline{j} \in I_A, |\underline{j}|=k+1} \int_{T_A v_{\underline{j}}(A)} \exp\left[\sum_{l=1}^k (\phi T^l - s_0)(v_{\underline{j}} x)\right] d\nu_A(x) + 1$$

because of the fact that $X_{\underline{i}} \subset TA(\forall \underline{i} \in I')$. By conformality of ν_A this coincides with

$$\sum_{k=1}^{\infty} \sum_{\underline{j} \in I_A, |\underline{j}|=k+1} \int_{v_{\underline{j}}(A)} \exp[(\sum_{l=1}^k \phi T^l - s_0)(x)] \exp[-\phi_A^{(s_0)}(x)] d\nu_A(x) + 1.$$

□

§7. THE CONSTRUCTION OF MUTUALLY SINGULAR EQUILIBRIUM STATES

In this section, we show the existence of mutually singular non-atomic equilibrium states by using induced systems.

Lemma 15. *Let (T, X, Q) be a transitive FRS Markov system and let $\phi \in \mathcal{W}(T)$. Suppose that there exists a sequence of full cylinders $\{X_i\}_{i=1}^M (M \leq \infty)$ that satisfies $(05)^*$ and $(06)^*$, $\inf_{x \in A_i} \phi(x) > -\infty$, and*

$$\|\mathcal{L}_{\phi_{A_i}^{\min\{0, P_{\text{top}}(T_{A_i}, \phi_{A_i}^{(0)})\}}} 1\| < \infty$$

for each $A_i = cl(int X_i)$. Let $\Gamma_0 := X$ and for each $i \geq 0$ define inductively $\Gamma_{i+1} = \bigcap_{n=0}^{\infty} T^{-n}(\Gamma_i \cap A_{i+1}^c) (\subset \Gamma_i)$. We assume that for each $i \geq 0$,

$$H_{\nu_{\Gamma_i \cap A_{i+1}}} (Q_{\Gamma_i \cap A_{i+1}}) < \infty, \quad \int_{\Gamma_i \cap A_{i+1}} R_{\Gamma_i \cap A_{i+1}} d\nu_{\Gamma_i \cap A_{i+1}} < \infty$$

for the Borel probability measure $\nu_{\Gamma_i \cap A_{i+1}}$ on $\Gamma_i \cap A_{i+1}$ obtained in Theorem 7. If $\bigcap_{i=1}^M \Gamma_i := \Gamma$ consists of periodic points, then there exists a T -invariant ergodic probability measure μ equivalent to an $\exp[P_{\text{top}}(T, \phi) - \phi]$ -conformal measure ν that satisfies the following variational principle:

$$P_{\text{top}}(T, \phi) = h_{\mu}(T) + \int_X \phi d\mu = \sup\{h_m(T) + \int_X \phi dm \mid m \in M_T(X, \phi) \text{ is ergodic}\}.$$

The equilibrium state μ for ϕ is not necessarily unique.

Theorem 8 (Phase transition and singular equilibrium states). *We assume all conditions in Lemma 15. If $\Gamma := \bigcap_{i=1}^M \Gamma_i$ consists of generalized indifferent periodic points with respect to ϕ , then there exists a sequence of ergodic equilibrium states $\{\mu_i\}_{i=1}^M$ that are mutually singular and the set of equilibrium states for ϕ is the convex hull of $\{\mu_i\}_{i=1}^M$ and the set of invariant Borel probability measures supported on Γ .*

Lemma 16 (Kac's formula). *If $\int_A R_A d(\nu|_A) < \infty$ and μ_A is a T_A -invariant ergodic probability measure equivalent to $\nu|_A$, then the next formula gives a T -invariant ergodic probability measure μ equivalent to ν :*

$$\mu(E)/\mu(A) = \int_A \sum_{i=0}^{R_A(x)-1} 1_E \circ T^i(x) d\mu_A(x) (\forall E \in \mathcal{F}).$$

Lemma 17 (Finite entropy condition). *Suppose that for $s_i := P_{\text{top}}(T|_{\Gamma_{i-1}}, \phi)$, $\int_{\Gamma_{i-1} \cap A_i} \phi_{\Gamma_{i-1} \cap A_i}^{(s_i)} d\nu_{\Gamma_{i-1} \cap A_i} > -\infty$. Then $H_{\nu_{\Gamma_i \cap A_{i+1}}} (Q_{\Gamma_i \cap A_{i+1}}) < \infty$.*

Proof of Lemma 17. Let $A = \Gamma_i \cap A_{i+1}$ and $s = s_i$. Since we have $\forall \underline{j} \in I_A$,

$$\nu_A(X_{\underline{j}}) = \int_A \exp[\phi_A^{(s)}(v_{\underline{j}}(x))] d\nu_A(x) \geq \exp[-s|\underline{j}|] \inf_{x \in v_{\underline{j}}(A)} \exp[\phi_A^{(0)}(x)],$$

the bounded distortion for $\phi_A^{(0)}$ allows us to see that

$$\begin{aligned} \int_A (-\phi_A^{(0)}) d\nu_A &\geq -\log C - \sum_{\underline{j} \in I_A} \int_{X_{\underline{j}}} \log(\inf_{x \in v_{\underline{j}}(A)} \exp[\phi_A^{(0)}(x)]) d\nu_A(x) \\ &\geq -\log C + \sum_{\underline{j} \in I_A} \int_{X_{\underline{j}}} (-s|\underline{j}|) d\nu_A - \sum_{\underline{j} \in I_A} \nu_A(X_{\underline{j}}) \log(\nu_A(X_{\underline{j}})), \end{aligned}$$

where C is the bounded distortion constant $\exp[D\gamma(1-\gamma)^{-1}(\text{diam } X)^\theta]$. These inequalities allow one to establish

$$H_{\nu_A}(Q_A) \leq \int_A (-\phi_A^{(0)}) d\nu_A + \log C + s \sum_{\underline{j} \in I_A} \int_{X_{\underline{j}}} R_A d\nu_A < \infty. \quad \square$$

Proof of Lemma 15. Since $\sum_{n=1}^{\infty} \text{Var}_n(T_{A_1}, \phi_{A_1}^{(s)}) < \infty$ is satisfied, it follows from Theorem 7 that $\exists \nu_{A_1}$ on A_1 satisfying $\mathcal{L}_{\phi_{A_1}^{(s)}}^* \nu_{A_1} = \nu_{A_1}$ for $s = P_{\text{top}}(T, \phi)$ and $\exists \nu$ an $\exp[P_{\text{top}}(T, \phi) - \phi]$ -conformal measure on X . Furthermore, by Proposition 5 we see that $\frac{\nu|_{A_1}}{\nu(A_1)} = \nu_{A_1}$. The bounded distortion allows one to obtain an ergodic T_{A_1} -invariant probability measure $\mu_{A_1} \sim \nu_{A_1}$ with a density $d\mu_{A_1}/d\nu_{A_1}$ away from zero and infinity. Furthermore, by [14] there exists an equilibrium state μ_{A_1} for $\phi_{A_1}^{(s)}$ with respect to T_{A_1} that is ergodic. In particular, since $H_{\nu_{A_1}}(Q_{A_1}) < \infty$ we have for $s_1 = P_{\text{top}}(T, \phi)$,

$$P_{\text{top}}(T_{A_1}, \phi_{A_1}^{(s_1)}) = 0 = h_{\mu_{A_1}}(T_{A_1}) + \int_{A_1} \phi_{A_1}^{(s_1)} d\mu_{A_1} \geq h_{m_{A_1}}(T_{A_1}) + \int_{A_1} \phi_{A_1}^{(s_1)} dm_{A_1}$$

for all T_{A_1} -invariant probability measures $m_{A_1} \in M_{T_{A_1}}(A_1, \phi_{A_1}^{(s_1)})$ (cf. [14]). These inequalities and Lemma 16 allow us to have a T -invariant ergodic probability measure $\mu_1 \sim \nu$ that satisfies $\mu_1(A_1) > 0$, $\mu_1(\Gamma_1) = 0$ and

$$\begin{aligned} (**) \quad 0 &= \mu_1(A_1)^{-1} (h_{\mu_1}(T) + \int_X (\phi - s_1) d\mu_1) \\ &\geq m(A_1)^{-1} (h_m(T) + \int_X (\phi - s_1) dm) \end{aligned}$$

for all T -invariant ergodic probability measures $m \in M_T(X, \phi)$ satisfying $m(A_1) > 0$. (**) is equivalent to the inequalities: $s_1 = h_{\mu_1}(T) + \int_X \phi d\mu_1 \geq h_m(T) + \int_X \phi dm$. On the other hand, any $m \in M_T(X, \phi)$ satisfying $m(A_1) = 0$ is supported on Γ_1 . In fact, since $X = (\bigcup_{i=0}^{\infty} T^{-i}A_1) \cup (\bigcap_{i=0}^{\infty} (T^{-i}A_1)^c)$, $m(A_1) = 0$ and T -invariance of m give $m(\Gamma_1) = 1$. Thus the set of all T -invariant probability measures m supported on Γ_1 coincides with the set of all T -invariant probability measures m with $m(A_1) = 0$. Since $(T|_{\Gamma_1}, \Gamma_1, Q_{\Gamma_1} := Q \cap \Gamma_1)$ is a subsystem of (T, X, Q) , we can apply the above arguments for the induced system $(T_{A_2 \cap \Gamma_1}, A_2 \cap \Gamma_1, Q_{A_2 \cap \Gamma_1})$. That is, for $s_2 = P_{\text{top}}(T|_{\Gamma_1}, \phi) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i: |i|=n} \sum_{v_i x = x, x \in \Gamma_1} \exp[\sum_{h=0}^{n-1} \phi T^h(x)]$ and for the associated potential $\phi_{A_2 \cap \Gamma_1}^{(s_2)}(x) = \sum_{h=0}^{R_{A_2 \cap \Gamma_1}(x)-1} \phi T^h(x) - s_2 R_{A_2 \cap \Gamma_1}(x)$

our assumptions allow us to establish the bounded distortion for $\phi_{A_2 \cap \Gamma_1}^{(s_2)}$ so that there exists a $T_{A_2 \cap \Gamma_1}$ -invariant ergodic probability measure $\mu_{A_2 \cap \Gamma_1}$ that satisfies

$$\begin{aligned} 0 &= h_{\mu_{A_2 \cap \Gamma_1}}(T_{A_2 \cap \Gamma_1}) + \int_{A_2 \cap \Gamma_1} \phi_{A_2 \cap \Gamma_1}^{(s_2)} d\mu_{A_2 \cap \Gamma_1} \\ &\geq h_{m_{A_2 \cap \Gamma_1}}(T_{A_2 \cap \Gamma_1}) + \int_{A_2 \cap \Gamma_1} \phi_{A_2 \cap \Gamma_1}^{(s_2)} dm_{A_2 \cap \Gamma_1} \end{aligned}$$

for all $T_{A_2 \cap \Gamma_1}$ -invariant ergodic probability measures $m_{A_2 \cap \Gamma_1} \in M_{T_{A_2 \cap \Gamma_1}}(A_2 \cap \Gamma_1, \phi_{A_2 \cap \Gamma_1}^{(s_2)})$. Let μ_2 be the T -invariant ergodic probability measure supported on Γ_1 arising from $\mu_{A_2 \cap \Gamma_1}$ via Kac's formula. Then $s_2 = h_{\mu_2}(T) + \int_X \phi d\mu_2 \geq h_m(T) + \int_X \phi dm$ for all T -invariant ergodic probability measures $m \in M_T(X, \phi)$ supported on Γ_1 that satisfy $m(\Gamma_1 \cap A_2) > 0$. Inductively we have a decreasing sequence $\{s_i\}_{i \in \mathbb{N}}$, where $s_i = P_{\text{top}}(T|_{\Gamma_{i-1}}, \phi)$ and a sequence of T -invariant ergodic probability measures $\{\mu_i\}_{i=1}^M$ such that μ_i is supported on Γ_{i-1} , $\mu_i(\Gamma_{i-1} \cap A_i) > 0$ and s_i satisfies $s_i = h_{\mu_i}(T) + \int_X \phi d\mu_i \geq h_m(T) + \int_X \phi dm$ for all T -invariant ergodic probability measures $m \in M_T(X, \phi)$ supported on Γ_{i-1} with $m(\Gamma_{i-1} \cap A_i) > 0$. Since $\mu_i(\Gamma_{i-1}) = 1$ and $\mu_i(\Gamma_i) = 0$, $\{\mu_i\}_{i=1}^M$ are mutually singular. Finally, for every T -invariant measure supported on Γ that consists of periodic points we have $s_1 \geq h_m(T) + \int_X \phi dm$. Since $\{s_i\}_{i \in \mathbb{N}}$ is decreasing, we complete the proof. \square

§8. APPLICATIONS

In this section, we show some examples of transitive FRS Markov systems to which our theorems 1-8 can apply.

Example 1 (Brun's map [13], [18], [20]). Let $X = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_2 \leq x_1 \leq 1\}$, and let

$$X_i = \{(x_1, x_2) \in X : x_i + x_1 \geq 1 \geq x_{i+1} + x_1\}$$

for $i = 0, 1, 2$ where we put $x_0 = 1$ and $x_3 = 0$. T is defined by

$$\begin{aligned} T(x_1, x_2) &= \left(\frac{x_1}{1-x_1}, \frac{x_2}{1-x_1}\right) \text{ on } X_0, \\ T(x_1, x_2) &= \left(\frac{1}{x_1} - 1, \frac{x_2}{x_1}\right) \text{ on } X_1, \\ T(x_1, x_2) &= \left(\frac{x_2}{x_1}, \frac{1}{x_1} - 1\right) \text{ on } X_2. \end{aligned}$$

Then $Q = \{X_i\}_{i=0}^2$ is a Bernoulli partition and $(0, 0)$ is an indifferent fixed point (i.e., $|\det DT(0, 0)| = 1$). Since T is a continuous piecewise C^2 map and $\sigma(n) = n^{-1}$, all conditions (01)-(04) are satisfied and dynamical instability is polynomial. We see that $\phi = -\log |\det DT|$ is piecewise Lipschitz continuous so that ϕ is a potential of WBV. Furthermore, since each periodic point is contained in a single cylinder the property (1) is satisfied. Define $B_1 = X_1 \cup X_2$. Then T^* satisfies the uniformly expanding property and a direct calculation allows us to establish (06) for $\phi = -\log |\det DT|$ (see [18] for more details). Hence we can apply Theorems 1-8. In particular, we can see that $P_{\text{top}}(T, \phi) = P_{\text{top}}(T^*, \phi^*) = 0$.

Example 2 (Inhomogeneous Diophantine approximations [13], [15], [16], [17], [19], [20], [21]). For the transformation defined in the introduction (Example A), we can directly verify all conditions (01)-(04). In fact, we can introduce an index set

$$I = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : a, b \in \mathbb{Z}, a > b > 0, \text{ or } a < b < 0 \right\}$$

and a partition $\left\{ X_{\begin{pmatrix} a \\ b \end{pmatrix}} : \begin{pmatrix} a \\ b \end{pmatrix} \in I \right\}$, where $X_{\begin{pmatrix} a \\ b \end{pmatrix}} = \{(x, y) \in X : a = \lfloor \frac{1-y}{x} \rfloor - \lfloor -\frac{y}{x} \rfloor, b = -\lfloor -\frac{y}{x} \rfloor\}$. Although $\phi = -\log |\det DT|$ fails (piecewise) Hölder continuity,

we can verify $V_n(\phi) \leq \log(1 + \sigma(n-2))$ and $\sigma(n) = O(n^{-1})$ (see [15], [20]) (cf. [16], [19], [21]). Hence $\phi = -\log|\det DT|$ is a potential of WBV. Since each periodic point is contained in a single cylinder, the property (1) in Lemma 6 is satisfied. Let D_n be the union of cylinders of rank n containing indifferent periodic points and let $B_n = D_{n-1} \setminus D_n$. Then the jump transformation $T^* : \bigcup_{i=1}^{\infty} B_i \rightarrow X$ defined by $T^*(x, y) = T^i(x, y)$ for $(x, y) \in B_i$ satisfies exponential decay of diameter of cylinders (see [19]), and we can verify the validity of (06) for $\phi = -\log|\det DT|$. Indeed, for $\underline{i} \in I^*$ with $|\underline{i}| = n$, $L_\phi(\underline{i}) \leq 3/n^2$ and so

$$\sup_{\underline{i} \in I^*} \sum_{j=0}^{|\underline{i}|-1} L_\phi(i_{j+1} \dots i_{|\underline{i}|}) \leq \sum_{n=1}^{\infty} \frac{3}{n^2} < \infty.$$

Hence we have summability of $\text{Var}_n(T^*, \phi^*)$, which allows us to apply Theorems 1 - 8.

Example 3 (A complex continued fraction [5], [12], [15], [22]). For the transformation T , defined in the introduction (Example B), we define $X_{n\alpha+m\bar{\alpha}} = \{z \in X : [1/z]_1 = n\alpha + m\bar{\alpha}\}$ for each $n\alpha + m\bar{\alpha} \in I := \{m\alpha + n\bar{\alpha} : (m, n) \in \mathbb{Z}^2 - (0, 0)\}$. Then we have a countable partition $Q = \{X_a\}_{a \in I}$ of X that is a topologically mixing Markov partition and satisfies (01)-(03). The inverse branches to T take the form $v_j(z) = 1/(j+z)$, where $j \in I$ and the v_j satisfy (04). Therefore the inverse branches of the n th iterate of the transformation T^n take the form

$$v_{j_1, \dots, j_n}(z) = \frac{p_n + zp_{n-1}}{q_n + zq_{n-1}} \text{ and } |v'_{j_1, \dots, j_n}(z)| = \frac{1}{|q_n + zq_{n-1}|^2}$$

where $p_n = j_n p_{n-1} + p_{n-2}$ and $q_n = j_n q_{n-1} + q_{n-2}$, $n \geq 1$, and $p_{-1} = \alpha$, $p_0 = 0 = q_{-1}$ and $q_0 = \alpha$. If the string j_1, \dots, j_{n-1} corresponds to a cylinder that contains one of the indifferent points, but the longer string j_1, \dots, j_n corresponds to a sub-cylinder disjoint from the indifferent periodic points, then v_{j_1, \dots, j_n} is an inverse branch of the jump transformation T^* which is uniformly expanding. For $\phi(z) = -\log|T'(z)|$, WBV and (06) are satisfied. Further details can be found in [22] in which multifractal formalism was established by applying our Theorems 1-8.

§9. APPENDIX - PROOFS OF RESULTS IN §§2 AND 3

For the proof of Theorem 1, we first verify the following facts.

Lemma 18. (18-1) $\forall U_k \in \mathcal{U}, \underline{Z}_n(U_k, \phi) > 0$ for all $n > S$.

(18-2) $\forall U_k \in \mathcal{U}, \forall n, m > S, \underline{Z}_{n+m}(U_k, \phi) \geq \underline{Z}_n(U_k, \phi) \underline{Z}_m(U_k, \phi)$. (Subadditivity)

(18-3) $\forall U_k, U_l \in \mathcal{U}$,

$$\begin{aligned} \underline{Z}_{n+s_l+s_k}(U_l, \phi) &\geq \underline{Z}_n(U_k, \phi) (C_{s_l} C_{s_k})^{-1} \\ &\times \sup_{x \in X^{(l,k)}(s_k)} \exp\left[\sum_{h=0}^{s_k-1} \phi T^h(x)\right] \sup_{x \in X^{(k,l)}(s_l)} \exp\left[\sum_{h=0}^{s_l-1} \phi T^h(x)\right]. \end{aligned}$$

Proof of Lemma 18. (18-1) follows from Lemma 1 and Remark (A). (18-2) follows from the definition of $\underline{Z}_n(U, \phi)$ directly. Since $\underline{Z}_{n+s_l+s_k}(U_l, \phi)$ is bounded from below by

$$\sum_{\underline{i}: |\underline{i}|=s_k, \text{int}(TX_{i_{s_k}})=U_k, \text{int}X_{i_1} \subset U_l} \inf_{x \in X_{\underline{i}}} \exp\left[\sum_{h=0}^{s_k-1} \phi T^h(x)\right]$$

$$\begin{aligned} & \times \sum_{\underline{j}: |\underline{j}|=n, \text{int}(TX_{j_n})=U_k \supset \text{int} X_{j_1}} \inf_{x \in X_{\underline{j}}} \exp \left[\sum_{h=0}^{n-1} \phi T^h(x) \right] \\ & \times \sum_{\underline{t}: |\underline{t}|=s_l, \text{int}(TX_{t_{s_l}})=U_l, \text{int} X_{t_1} \subset U_k} \inf_{x \in X_{\underline{t}}} \exp \left[\sum_{h=0}^{s_l-1} \phi T^h(x) \right] \end{aligned}$$

and the transitivity condition allows one to establish

$$\sum_{\underline{z}: |\underline{z}|=s_k, \text{int}(TX_{z_{s_k}})=U_k, \text{int} X_{z_1} \subset U_l} \inf_{x \in X_{\underline{z}}} \exp \left[\sum_{h=0}^{s_k-1} \phi T^h(x) \right] \geq \inf_{x \in X^{(l,k)}(s_k)} \exp \left[\sum_{h=0}^{s_k-1} \phi T^h(x) \right]$$

and

$$\sum_{\underline{t}: |\underline{t}|=s_l, \text{int}(TX_{t_{s_l}})=U_l, \text{int} X_{t_1} \subset U_k} \inf_{x \in X_{\underline{t}}} \exp \left[\sum_{h=0}^{s_l-1} \phi T^h(x) \right] \geq \inf_{x \in X^{(k,l)}(s_l)} \exp \left[\sum_{h=0}^{s_l-1} \phi T^h(x) \right],$$

(18-3) follows from the WBV property of ϕ . \square

Proof of Theorem 1. By (18-2) in Lemma 18 and the WBV property of ϕ , both $\lim_{n \rightarrow \infty} \frac{1}{n} \log \bar{Z}_n(U, \phi)$ and $\lim_{n \rightarrow \infty} \frac{1}{n} \log \underline{Z}_n(U, \phi)$ exist for each $U \in \mathcal{U}$. Since $\underline{Z}_n(U, \phi) \leq Z_n(U, \phi) \leq \bar{Z}_n(U, \phi)$, by the WBV property of ϕ , $\lim_{n \rightarrow \infty} \frac{1}{n} \log Z_n(U, \phi)$ also exists and all the limits coincide. By (18-3) in Lemma 18, it is obvious that the limit does not depend on U . Noting $\min_{1 \leq j \leq N} Z_n(\phi, U_j) \leq Z_n(\phi) \leq \sum_{j=1}^N Z_n(\phi, U_j)$ allows one to complete the proof. \square

In order to prove Theorem 3 we first show the next result.

Lemma 19. (19-1) $\forall V \in \mathcal{V}$ and $\forall x \in V, \mathcal{L}_\phi^n 1_V(x) \leq \sum_{U_l \supseteq V} \bar{Z}_n(U_l, \phi)$.

(19-2) $\forall x \in U_k \in \mathcal{U}, \mathcal{L}_\phi^n 1_{U_k}(x) \geq \underline{Z}_n(U_k, \phi)$.

(19-3) $Z_n(\phi) \geq C_1^{-1} C_{n-1}^{-1} \{ \min_{1 \leq k \leq N} \sup_{y \in X^{(k)}(1)} \exp \phi(y) \} \| \mathcal{L}_\phi^{n-2} 1 \|$, where $X^{(k)}(1) \in \mathcal{Q}$ satisfies $X^{(k)}(1) \subset U_k$ and $T(\text{int } X^{(k)}(1)) = \text{int } X$.

Proof of Lemma 19. We first note that $V = \bigcup_{X_j \subset V} X_j$ because of the Markov property of Q . Then for $x \in V$ the following inequalities allow us to have the assertion in (19-1):

$$\begin{aligned} \mathcal{L}_\phi^n 1_V(x) &= \sum_{j \in I: X_j \subset V} \mathcal{L}_\phi^n 1_{X_j}(x) = \sum_{j \in I: X_j \subset V} \sum_{(j i_2 \dots i_n): x \in TX_{i_n}} \exp \left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x) \right] \\ &= \sum_{j \in I: X_j \subset V} \sum_{U_l \in \mathcal{U}: x \in U_l} \sum_{(j i_2 \dots i_n): TX_{i_n} = U_l} \exp \left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x) \right] \\ &= \sum_{U_l \in \mathcal{U}: U_l \supset V} \sum_{(j i_2 \dots i_n): TX_{i_n} = U_l \supset V \supset X_j} \exp \left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x) \right] \\ &\leq \sum_{U_l \in \mathcal{U}: U_l \supset V} \sum_{(j i_2 \dots i_n): TX_{i_n} = U_l \supset X_j} \exp \left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x) \right] \\ &\leq \sum_{U_l \in \mathcal{U}: U_l \supset V} \bar{Z}_n(U_l, \phi). \end{aligned}$$

For $x \in U_k$, we have the following inequalities, which give (19-2):

$$\begin{aligned} \mathcal{L}_\phi^n 1_{U_k}(x) &= \sum_{j \in I: X_j \subset U_k} \sum_{(j i_2 \dots i_n): x \in TX_{i_n}} \exp\left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x)\right] \\ &= \sum_{j \in I: X_j \subset U_k} \sum_{U_l \in \mathcal{U}: x \in U_l} \sum_{(j i_2 \dots i_n): TX_{i_n} = U_l} \exp\left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x)\right] \\ &\geq \sum_{(j i_2 \dots i_n): TX_{i_n} = U_k \supset X_j} \exp\left[\sum_{h=0}^{n-1} \phi T^h(v_{j i_2 \dots i_n} x)\right] \geq \underline{Z}_n(U_k, \phi). \end{aligned}$$

By the WBV property of ϕ , $Z_n(\phi)$ is bounded from below by

$$C_1^{-1} C_{n-1}^{-1} \sum_{l=1}^N \sum_{\substack{\underline{i}: |\underline{i}|=n, \\ \text{int}(TX_{i_n})=U_l \supset \text{int}X_{i_1}}} \sup_{x \in X_{i_1}} \exp[\phi(x)] \sup_{x \in TX_{i_n}} \exp\left[\sum_{h=0}^{n-2} \phi T^h v_{i_2 \dots i_n}(x)\right].$$

Then (19-3) follows from the Markov property and the strong transitivity. □

Proof of Theorem 3. By (19-1,2) in Lemma 19, we have for $x \in V \subset U_k$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{L}_\phi^n 1_V(x) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log N\left(\max_{1 \leq l \leq N} \overline{Z}_n(U_l, \phi)\right) = P_{\text{top}}(T, \phi)$$

and $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{L}_\phi^n 1_{U_k}(x) \geq P_{\text{top}}(T, \phi)$. Our assumption $U_k \in \mathcal{V}$ implies $U_k = V$. Hence we have the first assertion. The rest of the assertions follow from (19-3) in Lemma 19 immediately. □

In order to prove Proposition 1, we need two sublemmas.

Sublemma A (Lemma 3.1 in [17]). $\bigcap_{n>0} D_n$ and X^* are positively T -invariant. Furthermore, $\bigcup_{m=1}^\infty T^{*-m}(\bigcap_{n \geq 0} D_n)$ contains no periodic points.

Proof. The result follows from the equality (4) : $R(Tx) = R(x) - 1$ ($R(x) \geq 2$). □

Sublemma B. Define

$$\begin{aligned} \mathcal{P}_n(X, T) &:= \{x \in X \mid \exists (i_1 \dots i_n) \in I^n \text{ such that } v_{i_1 \dots i_n} x = x\}; \\ \mathcal{P}_n(X^*, T^*) &:= \{x \in X \mid \exists (\underline{i}_1 \dots \underline{i}_n) \in I^{*n} \text{ such that } v_{\underline{i}_1 \dots \underline{i}_n} x = x\}. \end{aligned}$$

Then $\forall x \in \mathcal{P}_n(X, T) \cap X^*$, $\exists 0 < l \leq n$, $\exists y \in \mathcal{P}_l(X^*, T^*)$ such that $T^j y = x$ for some $j < R(y)$ and $\sum_{m=0}^{l-1} R(T^{*m}(y)) = n$. Furthermore, $\sum_{m=0}^{l-1} (\phi^* - sR)T^{*m}(y) = \sum_{h=0}^{n-1} (\phi - s)T^h(x)$. Conversely, $\forall y \in \mathcal{P}_l(X^*, T^*)$ and $\forall 0 \leq j < R(y)$, $T^j y \in \mathcal{P}_n(X, T) \cap X^*$ and $\sum_{m=0}^{l-1} R(T^{*m}(y)) = n$.

Proof. Since $x \in \mathcal{P}_n(X, T) \cap X^*$ visits B_1 infinitely often, we can find a point $y \in \mathcal{P}_l(X^*, T^*)$ for some $l \leq n$ such that $T^j y = x$ for some $j < R(y)$ and

$\sum_{m=0}^{l-1} R(T^{*m}(y)) = n$. By the property (4) the converse is also true. Since

$$\begin{aligned} \sum_{m=0}^{l-1} (\phi^* - sR)T^{*m}(y) &= \sum_{m=0}^{l-1} \sum_{h=0}^{R(T^{*m}y)-1} \phi T^h(T^{*m}y) - sn \\ &= \sum_{h=0}^{\sum_{m=0}^{l-1} R(T^{*m}y)-1} (\phi - s)T^h(y) \end{aligned}$$

we have the rest of the assertion. \square

Proof of Proposition 1. By Sublemma A we first note that $\mathcal{P}_n(X, T) = \{\mathcal{P}_n(X, T) \cap \bigcap_{n \geq 0} D_n\} \cup \{\mathcal{P}_n(X, T) \cap X^*\}$. Then we see that $\zeta_{T, \phi}(\exp(-s))$ is equal to

$$\zeta_{T|_{\bigcap_{n \geq 0} D_n}, \phi}(\exp(-s)) \exp\left[\sum_{n=1}^{\infty} \frac{\exp[-ns]}{n} \times \sum_{x \in \mathcal{P}_n(X, T) \cap X^*} \exp\left[\sum_{h=0}^{n-1} \phi T^h(x)\right]\right].$$

Define for $n \geq l > 0$ $E_{n,l}^* := \{y \in \mathcal{P}_l(X^*, T^*) : \sum_{m=0}^{l-1} R(T^{*m}(y)) = n\}$ and

$$E_{n,l} := \{x \in X^* \cap \mathcal{P}_n(X, T) : \exists y \in X^* \cap \mathcal{P}_l(X^*, T^*) \text{ and } \exists j < R(y)$$

$$\text{such that } \sum_{m=0}^{l-1} R(T^{*m}(y)) = n \text{ and } x = T^j y\}.$$

Then $\mathcal{P}_l(X^*, T^*) = \bigcup_{n \geq l} E_{n,l}^*$ and $\mathcal{P}_n(X, T) \cap X^* = \bigcup_{l \leq n} E_{n,l}$. By Sublemma B we see that for $x \in X^*$,

$$\begin{aligned} &\sum_{n=1}^{\infty} \frac{1}{n} \sum_{x \in \mathcal{P}_n(X, T) \cap X^*} \exp\left[\sum_{h=0}^{n-1} (\phi - s)T^h(x)\right] \\ &= \sum_{n=1}^{\infty} \sum_{l \leq n} \sum_{\{x, Tx, \dots, T^{n-1}x\} \subset E_{n,l}} \exp\left[\sum_{h=0}^{n-1} (\phi - s)T^h(x)\right] \\ &= \sum_{n=1}^{\infty} \sum_{l \leq n} \sum_{\{y, T^*y, \dots, T^{*l-1}y\} \subset E_{n,l}^*} \exp\left[\sum_{m=0}^{l-1} (\phi^* - sR)T^{*m}(y)\right] \\ &= \sum_{l=1}^{\infty} \sum_{\{y, T^*y, \dots, T^{*l-1}y\} \subset \mathcal{P}_l(X^*, T^*)} \exp\left[\sum_{m=0}^{l-1} (\phi^* - sR)T^{*m}(y)\right] = \zeta_{T^*, \phi^* - sR}(1). \end{aligned}$$

We complete the proof. \square

Proof of Lemma 7. (i) By Lemmas 13-14, we have continuity of the function $P_{\text{top}}(T^*, \phi^* - sR)$ on $\text{int}\{s \in \mathbb{R} \mid P_{\text{top}}(T^*, \phi^*) \in \mathbb{R}\}$. Then the existence of a zero $s_0 \geq 0$ of (GBE) follows from the standard argument. Since $P_{\text{top}}(T^*, \phi^* - s_0R) = 0$, by Corollary 1 we have $s_0 \leq P_{\text{top}}(T, \phi)$. If $P_{\text{top}}(T, \phi) < 0$, then we have a contradiction. For (ii), replacing ϕ by $\phi - P_{\text{top}}(T^*, \phi^*)$ allows us to reduce to the case (i). (iii) Since the case $0 \leq P_{\text{top}}(T^*, \phi^*) < \infty$ is covered by Lemma 7(i), we suppose either $P_{\text{top}}(T^*, \phi^*) = \infty$ or $P_{\text{top}}(T^*, \phi^*) < 0$. If $P_{\text{top}}(T^*, \phi^*) = \infty$, then $\sup\{s \in \mathbb{R} : P_{\text{top}}(T^*, \phi^* - sR) = \infty\} = 0$. Hence $\forall s > 0$, $P_{\text{top}}(T^*, \phi^* - sR) < \infty$. Since the function $s \rightarrow P_{\text{top}}(T^*, \phi^* - sR)$ is decreasing and continuous on $\text{int}\{s \in \mathbb{R} \mid P_{\text{top}}(T^*, \phi^* - sR) \in \mathbb{R}\}$, we have $\lim_{s \rightarrow 0} P_{\text{top}}(T^*, \phi^* - sR) = \infty$ so that for sufficiently small $s > 0$, $P_{\text{top}}(T^*, \phi^* - sR) > 0$. On the other hand, it follows from

Corollary 1 that $P_{\text{top}}(T^*, \phi^* - sR) \leq 0$ for $s > P_{\text{top}}(T, \phi)$. Hence we have a zero $s_0 \geq 0$ of (GBE). If $P_{\text{top}}(T^*, \phi^*) < 0$, then $\sup\{s \in \mathbb{R} : P_{\text{top}}(T^*, \phi^* - sR) = \infty\} = P_{\text{top}}(T^*, \phi^*)$. The same argument as those for the previous case allows us to have a zero $s_0 \geq P_{\text{top}}(T^*, \phi^*)$ of (GBE). \square

Proof of Theorem 4. By Lemma 8 we have for $\underline{i} \in I^*$, $\frac{d(\nu v_{\underline{i}})}{d\nu} = \exp[\phi^* - s_0 R]v_{\underline{i}}$. If $\underline{i} = i_1 i_2 \dots i_n$, then $\frac{d(\nu v_{\underline{i}})}{d\nu}(x) = \exp[\sum_{h=0}^{n-1} \phi T^h v_{i_1 i_2 \dots i_n} x - s_0 n]$. Since the property (4) : $R(Tx) = R(x) - 1$ ($R(x) \geq 2$) implies $i_2 i_3 \dots i_n \in I^*$, the equality

$$\frac{d(\nu v_{\underline{i}})}{d\nu} = \frac{d(\nu v_{i_1})}{d\nu}(i_2 i_3 \dots i_n x) \frac{d(\nu v_{i_2 i_3 \dots i_n})}{d\nu}(x)$$

allows one to see that $\forall X_i \subset D_1$, $\frac{d(\nu v_{\underline{i}})}{d\nu}(x) = \exp[\phi v_i(x) - s_0](\forall x \in X^*)$. On the other hand, we know that the above equality holds for $X_i \subset B_1$ since $i \in I^*$. Finally, we establish $\forall X_i \in Q$, $\frac{d(\nu T)}{d\nu}|_{X_i}(x) = \exp[s_0 - \phi(x)](\forall x \in X^*)$. It follows from Lemma 2 and Theorem 2.1 in [18] that $s_0 = P_{\text{top}}(T, \phi)$. The assertion $\nu(\bigcup_{i \in I} \partial X_i) = \nu(\bigcup_{i \in I} \partial v_i(X)) = 0$ follows from $\nu(\text{int} X) = 1$, which is obtained by Lemma 8. We complete the proof. \square

REFERENCES

1. J. Aaronson, M. Denker and M. Urbański, *Ergodic theory for Markov fibred systems and parabolic rational maps*, Trans. Amer. Math. Soc. **337** (1993), 495-548. MR **94g**:58116
2. R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Mathematics 470, Springer, 1975. MR **56**:1364
3. M. Denker, F. Przytycki and M. Urbański, *On the transfer operator for rational functions on the Riemann sphere*, Ergodic Theory and Dynamical Systems **16** (1996), 255-266. MR **97e**:58197
4. M. Denker and M. Urbański, *On the existence of conformal measures*, Trans. Amer. Math. Soc. **328** (1991), 563-587. MR **92k**:58155
5. M. Denker and M. Yuri, *A note on the construction of nonsingular Gibbs measures*, Colloquium Mathematicum **84/85** (2000), 377-383. MR **2001k**:37013
6. M. Denker and M. Yuri, *Partially defined infinite iterated functional systems*, In preparation.
7. D. Fiebig, U. Fiebig and M. Yuri, *Pressures and Equilibrium states for countable Markov shifts*, To appear in Israel J. Math.
8. P. H. Hanus, R. D. Mauldin and M. Urbański, *Thermodynamic formalism and multi-fractal analysis of conformal infinite iterated functional systems*, Acta Math. Hungar. **96** (2002), 27-98.
9. R. D. Mauldin and M. Urbański, *Parabolic iterated function systems*, Ergodic Theory and Dynamical Systems **20** (2000), 1423-1447. MR **2001m**:37047
10. Omri Sarig, *Thermodynamic formalism for countable Markov shifts.*, Ergodic Theory and Dynamical Systems **19** (1999), 1565-1593. MR **2000m**:37009
11. Y. B. Pesin, *Dimension Theory in Dynamical Systems. Contemporary Views and Applications*, Chicago University Press, 1997. MR **99b**:58003
12. M. Pollicott and M. Yuri, *Zeta functions for certain multi-dimensional nonhyperbolic maps*, Nonlinearity **14** (2001), 1265-1278. MR **2002h**:37036
13. F. Schweiger, *Ergodic theory of fibred systems and metric number theory*, Oxford University Press, Oxford, 1995. MR **97h**:11083
14. P. Walters, *Invariant measures and equilibrium states for some mappings which expand distances*, Trans. Amer. Math. Soc. **236** (1978), 121-153. MR **57**:6371
15. M. Yuri, *On a Bernoulli property for multi-dimensional mappings with finite range structure*, Tokyo J. Math **9** (1986), 457-485. MR **88d**:28023
16. M. Yuri, *On the convergence to equilibrium states for certain nonhyperbolic systems*, Ergodic Theory and Dynamical Systems **17** (1997), 977-1000. MR **98f**:58155
17. M. Yuri, *Zeta functions for certain nonhyperbolic systems and topological Markov approximations*, Ergodic Theory and Dynamical Systems **18** (1998), 1589-1612. MR **2000j**:37024

18. M. Yuri, *Thermodynamic formalism for certain nonhyperbolic maps*, Ergodic Theory and Dynamical Systems **19** (1999), 1365-1378. MR **2001a**:37012
19. M. Yuri, *Statistical properties for nonhyperbolic maps with finite range structure*, Trans. Amer. Math. Soc. **352** (2000), 2369-2388. MR **2000j**:37009
20. M. Yuri, *Weak Gibbs measures for certain nonhyperbolic systems*, Ergodic Theory and Dynamical Systems **20** (2000), 1495-1518. MR **2002d**:37011
21. M. Yuri, *On the speed of convergence to equilibrium states for multi-dimensional maps with indifferent periodic points.*, Nonlinearity **15** (2002), 429-445. MR **2002k**:37006
22. M. Yuri, *Multifractal Analysis of weak Gibbs measures for Intermittent Systems.*, Commun. Math. Phys. **230** (2002), 365-388.

DEPARTMENT OF BUSINESS ADMINISTRATION, SAPPORO UNIVERSITY, NISHIOKA, TOYOHIRA-KU,
SAPPORO 062-8520, JAPAN

E-mail address: yuri@math.sci.hokudai.ac.jp, yuri@mail-ext.sapporo-u.ac.jp

STRONGLY INDEFINITE FUNCTIONALS AND MULTIPLE SOLUTIONS OF ELLIPTIC SYSTEMS

D. G. DE FIGUEIREDO AND Y. H. DING

ABSTRACT. We study existence and multiplicity of solutions of the elliptic system

$$\begin{cases} -\Delta u = H_u(x, u, v) & \text{in } \Omega, \\ -\Delta v = -H_v(x, u, v) & \text{in } \Omega, \quad u(x) = v(x) = 0 \quad \text{on } \partial\Omega, \end{cases}$$

where $\Omega \subset \mathbb{R}^N$, $N \geq 3$, is a smooth bounded domain and $H \in C^1(\bar{\Omega} \times \mathbb{R}^2, \mathbb{R})$. We assume that the nonlinear term

$$H(x, u, v) \sim |u|^p + |v|^q + R(x, u, v) \quad \text{with} \quad \lim_{|(u,v)| \rightarrow \infty} \frac{R(x, u, v)}{|u|^p + |v|^q} = 0,$$

where $p \in (1, 2^*)$, $2^* := 2N/(N - 2)$, and $q \in (1, \infty)$. So some supercritical systems are included. Nontrivial solutions are obtained. When $H(x, u, v)$ is even in (u, v) , we show that the system possesses a sequence of solutions associated with a sequence of positive energies (resp. negative energies) going toward infinity (resp. zero) if $p > 2$ (resp. $p < 2$). All results are proved using variational methods. Some new critical point theorems for strongly indefinite functionals are proved.

1. INTRODUCTION AND MAIN RESULTS

Consider the following elliptic system:

$$(E) \quad \begin{cases} -\Delta u = H_u(x, u, v) & \text{in } \Omega, \\ -\Delta v = -H_v(x, u, v) & \text{in } \Omega, \\ u(x) = v(x) = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\Omega \subset \mathbb{R}^N$, $N \geq 3$, is a smooth bounded domain and $H : \bar{\Omega} \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is a C^1 -function. Here H_u denotes the partial derivative of H with respect to the variable u . Writing $z := (u, v)$, we suppose $H(x, 0) \equiv 0$ and $H_z(x, 0) \equiv 0$. Then $z = 0$ is a trivial solution of the system. In this paper we discuss the existence of nontrivial solutions. Roughly speaking, we are mainly interested in the class of Hamiltonians H such that

$$H(x, u, v) \sim |u|^p + |v|^q + R(x, u, v) \quad \text{with} \quad \lim_{|z| \rightarrow \infty} \frac{R(x, u, v)}{|u|^p + |v|^q} = 0,$$

where $1 < p < 2^* := 2N/(N - 2)$ and $q > 1$. The most interesting results obtained here refer to the case when $q \geq 2^*$, which correspond to critical and supercritical problems. The case when $q < 2^*$ has been studied by Costa and Magalhães [5],

Received by the editors June 18, 2001.

2000 *Mathematics Subject Classification*. Primary 35J50; Secondary 58E99.

Key words and phrases. Elliptic system, multiple solutions, critical point theory.

[6] and Benci and Rabinowitz [3]. See also Bartsch and De Figueiredo [2], De Figueiredo and Magalhães [7], De Figueiredo and Felmer [8] and Hulshof and van der Vorst [11], where similar systems also leading to strongly indefinite functionals have been studied. However, only subcritical systems have been considered in those papers.

Letting $2_* = 2^*/(2^* - 1) = 2N/(N + 2)$, we assume that $H(x, z)$ satisfies the following condition:

(H_0) there are $p \in (1, 2^*)$, $q \in (1, \infty)$ and $\tau \in (1, 1 + q/2_*)$ such that, for all (x, z) ,

$$|H_u(x, u, v)| \leq \gamma_0(1 + |u|^{p-1} + |v|^{\tau-1})$$

and

$$|H_v(x, u, v)| \leq \gamma_0(1 + |u|^{p-1} + |v|^{q-1}).$$

In all hypotheses on $H(x, z)$ the γ_i 's denote positive constants independent of (x, z) . We note that if $q < 2^*$, then $2_* < q/(q - 1)$, i.e., $q - 1 < q/2_*$. Hence, it is possible that $q \leq \tau < 1 + q/2_*$. However, if $q \geq 2^*$, then $\tau < q$. Furthermore, we remark that τ can be very large, if q is sufficiently large.

In addition, we need distinct conditions on H corresponding to the cases when $p > 2$, $p < 2$ or $p = 2$.

First, consider the case when $p > 2$. In this case, we assume the following three conditions:

(H_1) there are $\mu > 2$, $\nu > 1$ and $R_1 \geq 0$ such that

$$\frac{1}{\mu}H_u(x, z)u + \frac{1}{\nu}H_v(x, z)v \geq H(x, z) \quad \text{whenever } |z| \geq R_1,$$

with the provision that $\nu = \mu$ if $q > 2$;

(H_2) there are $2_*(p - 1) \leq \alpha \leq p$ and $2_*(\tau - 1) < \beta$ such that

$$H(x, z) \geq \gamma_1(|u|^\alpha + |v|^\beta) - \gamma_2 \quad \text{for all } (x, z),$$

and $\beta = q$ if $q > 2^*$;

(H_3) $H(x, 0, v) \geq 0$ and $H_u(x, u, 0) = o(|u|)$ as $u \rightarrow 0$ uniformly in x .

We prove the following results.

Theorem 1.1. *Let (H_0) be satisfied with $p > 2$. If $(H_1) - (H_3)$ hold, then (E) has at least one nontrivial solution.*

In order to provide some more transparent hypotheses under which the above result holds, we next present some conditions on H that are sufficient for (H_0) , (H_1) and (H_2) to hold:

(H'_0) there are $p \in (1, 2^*)$ and $q \in (2, \infty)$ such that, for all (x, z) ,

$$|H_u(x, u, v)| \leq \gamma_0(1 + |u|^{p-1} + |v|^{\frac{q}{2}-1})$$

and

$$|H_v(x, u, v)| \leq \gamma_0(1 + |u|^{p-1} + |v|^{q-1});$$

(H'_1) there are $\mu > 2$ and $R_1 \geq 0$ such that

$$H_u(x, z)u + H_v(x, z)v \geq \mu H(x, z) \quad \text{whenever } |z| \geq R_1;$$

(H'_2) for p and q as above,

$$H(x, z) \geq \gamma_1(|u|^p + |v|^q) - \gamma_2 \quad \text{for all } (x, z).$$

Theorem 1.1'. Let (H'_0) be satisfied with $p > 2$. If (H'_1) , (H'_2) , and (H_3) hold, then (E) has at least one nontrivial solution.

Theorem 1.2. Let (H_0) be satisfied with $p > 2$. If $H(x, z)$ is even in z and satisfies (H_1) and (H_2) , then (E) has a sequence (z_n) of solutions with energies $I(z_n) := \int_{\Omega} \left(\frac{1}{2}(|\nabla u_n|^2 - |\nabla v_n|^2) - H(x, z_n) \right)$, going to ∞ as $n \rightarrow \infty$.

In order to describe the other results, let $\sigma(-\Delta)$ denote the set of all eigenvalues of $(-\Delta, H_0^1(\Omega))$: $\lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots$.

We now consider the case when $p < 2$. We make the following assumptions:

(H_4) there are $\mu \in (1, 2)$, $\nu \geq 2$ and $\gamma_3 \geq 0$ ($\gamma_3 = 0$, if $q > 2^*$) such that

$$H(x, u, v) \geq \frac{1}{\mu} H_u(x, u, v)u + \frac{1}{\nu} H_v(x, u, v)v - \gamma_3 \quad \text{for all } (x, z);$$

(H_5) there are $\alpha \in (1, 2)$ and $\delta \in (0, 1/2)$ such that $H(x, u, v) \geq \gamma_4 |u|^\alpha - \delta \lambda_1 v^2$ for all (x, z) ;

(H_6) if $q \geq 2^*$, then $H_v(x, z)v \geq \gamma_5 |v|^q - \gamma_6 (|v| + u^2)$ for all (x, z) .

With these assumptions we have the following three results, for the case when $p < 2$.

Theorem 1.3. Suppose that (H_0) holds with $p < 2$ and $q \geq 2$. If $H(x, z)$ also satisfies $(H_4) - (H_6)$, then (E) has at least one nontrivial solution.

Theorem 1.4. Suppose that $H(x, z)$ is even in z and (H_0) holds with $p < 2$ and $q \geq 2$. If $H(x, z)$ also satisfies $(H_4) - (H_6)$, then (E) has a sequence (z_n) of solutions with negative energies $I(z_n)$ going to 0 as $n \rightarrow \infty$.

Theorem 1.5. Let (H_0) , with $p, q \in (1, 2)$, and (H_5) be satisfied. Then (E) has at least one nontrivial solution. If, in addition, $H(x, z)$ is even in z , then (E) has a sequence (z_n) of solutions with negative energies $I(z_n)$ going to 0 as $n \rightarrow \infty$.

Finally, we consider the case when $p = 2$, which presents some sort of resonance. Assume

(H_7) there exist $b_0 \leq 0 < a_0$ such that $R_0(x, z) := H(x, z) - \frac{1}{2}(a_0 u^2 + b_0 v^2) = o(|z|^2)$ as $z \rightarrow 0$ uniformly in x ;

(H_8) there exist $\sigma \in (1, 2)$, $a_\infty \in [a_0, \infty) \setminus \sigma(-\Delta)$, such that $R_\infty(x, z) := H(x, z) - \frac{1}{2}a_\infty u^2$ satisfies $|\partial_u R_\infty(x, z)| \leq \gamma_7(1 + |u|^{\sigma-1} + |v|^{\tau-1})$ and $R_\infty(x, z) \geq \gamma_8 |v|^q - \gamma_9(1 + |u|^\sigma)$.

The position of the numbers a_0, a_∞, b_0 with respect to the spectrum $\sigma(-\Delta)$ plays a very essential role in the next result. For that matter, let i, j, k be nonnegative integers such that $\lambda_i = \min\{\lambda \in \sigma(-\Delta) : \lambda > a_0\}$, $\lambda_j = \max\{\lambda \in \sigma(-\Delta) : \lambda < -b_0\}$, $\lambda_k = \max\{\lambda \in \sigma(-\Delta) : \lambda < a_\infty\}$, and set

$$\ell = \begin{cases} j & \text{if } a_\infty = a_0, \\ j + k - i + 1 & \text{if } a_\infty > a_0. \end{cases}$$

Now we can state our last result.

Theorem 1.6. Let (H_0) be satisfied with $p = 2$ and $\tau < 1 + q/2$. Assume that $H(x, z)$ is even in z and satisfies (H_7) and (H_8) . Then (E) has at least one pair of nontrivial solutions if $\ell = 1$, and infinitely many solutions if $\ell \geq 2$.

The cases covered in Theorem 1.6 include some asymptotically linear systems. Such systems have been studied in [5], [6] and Silva [13]. However, their results are not comparable with the ones obtained here.

We organize the paper as follows. In order to establish multiplicity of solutions we need some new abstract propositions on critical point theory for strongly indefinite functionals, which will be provided in Section 2. These propositions are based on certain Galerkin approximations, and we emphasize that the functionals do not satisfy the usual Palais-Smale condition. In Section 3 we study systems that are superlinear in the variable u , and prove Theorems 1.1 and 1.2. In Section 4 we consider systems that are sublinear in the variable u , and prove Theorems 1.3, 1.4 and 1.5. In both Sections 3 and 4, the variable v can have subcritical growth as well as supercritical growth. Finally, in Section 5, we consider a special asymptotically linear system and prove existence of multiple solutions.

2. CRITICAL POINTS FOR STRONGLY INDEFINITE FUNCTIONALS

Let E be a Banach space with norm $\|\cdot\|$. Suppose that E has a direct sum decomposition $E = E^1 \oplus E^2$ with both E^1 and E^2 being infinite dimensional. Let P^1 denote the projection from E onto E^1 . Assume (e_n^1) (resp. (e_n^2)) is a basis for E^1 (resp. E^2). Set

$$X_n := \text{span}\{e_1^1, \dots, e_n^1\} \oplus E^2, \quad X^m := E^1 \oplus \text{span}\{e_1^2, \dots, e_m^2\},$$

and let $(X^m)^\perp$ denote the complement of X^m in E . For a functional $I \in C^1(E, \mathbb{R})$ we set $I_n := I|_{X_n}$, the restriction of I on X_n . Recall that a sequence $(z_j) \subset E$ is said to be a $(\text{PS})_c^*$ sequence if $z_j \in X_{n_j}$, $n_j \rightarrow \infty$, $I(z_j) \rightarrow c$ and $I'_{n_j}(z_j) \rightarrow 0$ as $j \rightarrow \infty$. If any $(\text{PS})_c^*$ sequence has a convergent subsequence, then we say that I satisfies the $(\text{PS})_c^*$ condition.

Denote the upper and lower level sets, respectively, by $I_a = \{z \in E : I(z) \geq a\}$, $I^b = \{z \in E : I(z) \leq b\}$ and $I_a^b = I_a \cap I^b$ (denote similarly $(I_n)_a$, $(I_n)^b$ and $(I_n)_a^b$). We also set $\mathcal{K} = \{z \in E : I'(z) = 0\}$, $\mathcal{K}_c = \mathcal{K} \cap I_c$, $\mathcal{K}^c = \mathcal{K} \cap I^c$ and $\mathcal{K}_a^b = \mathcal{K}_a \cap \mathcal{K}^b$.

Proposition 2.1. *Let E be as above and let $I \in C^1(E, \mathbb{R})$ be even with $I(0) = 0$. In addition, suppose that, for each $m \in \mathbb{N}$, the conditions below hold:*

- (I₁) *there is $R_m > 0$ such that $I(z) \leq 0$ for all $z \in X^m$ with $\|z\| \geq R_m$;*
- (I₂) *there are $r_m > 0$ and $a_m \rightarrow \infty$ such that $I(z) \geq a_m$ for all $z \in (X^{m-1})^\perp$ with $\|z\| = r_m$;*
- (I₃) *I is bounded from above on bounded sets of X^m ;*
- (I₄) *if $c \geq 0$, any $(\text{PS})_c^*$ sequence (z_n) has a subsequence along which $z_n \rightarrow z \in \mathcal{K}_c$.*

Then the functional I has a sequence (c_k) of critical values, with the property that $c_k \rightarrow \infty$.

Remark 2.1. This proposition is more or less known if the condition (I₄) is replaced by the $(\text{PS})^*$ condition (cf. [1], [9]), or by the usual Palais-Smale condition, that is, any sequence $(z_k) \subset E$ such that $|I(z_k)| \leq c$ and $I'(z_k) \rightarrow 0$ has a convergent subsequence (cf. [3]).

Proposition 2.2. *Let E be as above and let $I \in C^1(E, \mathbb{R})$ be even. Assume that $I(0) = 0$ and that, for each $m \in \mathbb{N}$, the two conditions below hold:*

- (I₅) *there are $r_m > 0$ and $a_m > 0$ such that $I(z) \geq a_m$ for all $z \in X^m$ with $\|z\| = r_m$;*
- (I₆) *there is $b_m > 0$ with $b_m \rightarrow 0$ such that $I(z) \leq b_m$ for all $z \in (X^{m-1})^\perp$.*

Moreover, suppose that either I satisfies the $(PS)_c^*$ condition for all $c > 0$, or that the condition below holds:

(I_7) $\inf I(K) = 0$, and, for all $c \geq 0$, any $(PS)_c^*$ sequence (z_n) has a subsequence along which $z_n \rightarrow z \in K^c$ with $z = 0$ only if $c = 0$.

Then I has a sequence (c_k) of positive critical values satisfying $c_k \rightarrow 0$.

Proof. Let Σ be the family of symmetric, closed subsets of $E \setminus \{0\}$, and let $\gamma : \Sigma \rightarrow \mathbb{N} \cup \{0, \infty\}$ denote the Krasnoselski genus map. Set

$$c_n^m := \sup_{A \in \Sigma_n^m} \inf_{z \in A} I(z),$$

where

$$\Sigma_n^m := \{A \in \Sigma : A \subset X_n \text{ and } \gamma(A) \geq n + m\}.$$

Fix $m \in \mathbb{N}$. The Borsuk-Ulam theorem implies that $A \cap (X^{m-1})^\perp \neq \emptyset$ for each $A \in \Sigma_n^m$. It follows from (I_6) that

$$\inf_{z \in A} I(z) \leq \sup_{z \in (X^{m-1})^\perp} I(z) \leq b_m.$$

On the other hand, since $\gamma(\partial B_{r_m} \cap X_n^m) = n + m$, one has $S_n^m := \partial B_{r_m} \cap X_n^m \in \Sigma_n^m$, and so, by (I_5) , we obtain

$$\inf_{z \in S_n^m} I(z) \geq a_m.$$

Therefore,

$$(2.1) \quad a_m \leq c_n^m \leq b_m.$$

A standard deformation argument, using a positive pseudo-gradient flow, yields the existence of a sequence $(z_n^m)_{n=1}^\infty$, with $z_n^m \in X_n$ satisfying

$$|I(z_n^m) - c_n^m| \leq \frac{1}{n} \quad \text{and} \quad \|I'_n(z_n^m)\| \leq \frac{1}{n}.$$

We can assume that $I(z_n^m) \rightarrow c_m$ as $n \rightarrow \infty$. So, (z_n^m) is a $(PS)_{c_m}^*$ sequence with

$$(2.2) \quad a_m \leq c_m \leq b_m.$$

Now, if we assume that I satisfies the $(PS)_c^*$ condition for $c > 0$, then the conclusion follows. Next, suppose instead that (I_7) holds. Then, along a subsequence, $z_n^m \rightarrow z_m$ as $n \rightarrow \infty$ with $I'(z_m) = 0$ and $0 < I(z_m) \leq c_m$. Finally, by (2.2),

$$I(z_m) \leq b_m \rightarrow 0,$$

and the proof is complete. \square

Proposition 2.3. *Let E be as above and let $I \in C^1(E, \mathbb{R})$ be even with $I(0) = 0$. Suppose, in addition, that the three conditions below hold:*

- (I_8) *there are $\ell \in \mathbb{N}$ and $r, a > 0$ such that $I(z) \geq a$ for all $z \in X^\ell$ with $\|z\| = r$;*
- (I_9) *there is $b > 0$ such that $\sup I(E^2) \leq b$;*
- (I_{10}) *any $(PS)_c^*$, $c > 0$, sequence (z_n) has a subsequence along which $z_n \rightarrow z \in K_c$ and $P^1 z_n \rightarrow P^1 z$.*

Then I has at least one pair of nontrivial critical points if $\ell = 1$, and infinitely many critical points if $\ell > 1$, with positive critical values.

Proof. Let Σ , γ , Σ_n^m and c_n^m be as in the proof of Proposition 2.2. As before, by (I_8) and (I_9) , we obtain

$$a \leq c_n^m \leq b \quad \text{for all } n \in \mathbb{N} \text{ and } m = 1, \dots, \ell,$$

and we find sequences $z_n^m \in X_n$ such that, going to subsequences if necessary, $I(z_n^m) \rightarrow c_m$ and $I'_n(z_n^m) \rightarrow 0$ as $n \rightarrow \infty$ with

$$b \geq c_1 \geq c_2 \geq \dots \geq c_\ell \geq a.$$

Using (I_{10}) , we can assume furthermore that $z_n^m \rightarrow z_m \in \mathcal{K}_{c_m}$ for $m = 1, \dots, \ell$, as $n \rightarrow \infty$. If $\ell = 1$ the proof is complete.

Consider $\ell > 1$. Let $F = \{z \in \mathcal{K} : I(z) > 0\}$. We are going to prove that F is an infinite set. Arguing by contradiction, we suppose that F is finite. Choose $0 < \mu < a \leq b < \nu$ satisfying

$$\mu < \inf I(F) \leq \sup I(F) < \nu.$$

Let $k \in \mathbb{N}$ be so large that $0 \notin A := Q^k F$, where $Q^k : E \rightarrow X^k$ denotes the projection. Then A is also finite, and $\gamma(A) = 1$. By the continuity of γ , for all $\delta > 0$ small, $\gamma(N_\delta^k(A)) = \gamma(A)$, where $N_\delta^k(A) = \{z \in X^k : \text{dist}(z, A) \leq \delta\}$. Set $C_\delta = N_\delta^k(A) \oplus (X^k)^\perp$. Since $N_\delta^k(A) \subset C_\delta$ and $Q^k : C_\delta \rightarrow N_\delta^k(A)$, it follows from the properties of γ that $\gamma(C_\delta) = \gamma(N_\delta^k(A))$. We remark that $Q^k = P^1 + (Q^k - P^1)$ and that the range of $Q^k - P^1$ is k -dimensional. So by virtue of (I_{10}) , we conclude that, for all $c \geq 0$, any $(\text{PS})_c^*$ sequence (z_n) has a subsequence along which $z_n \rightarrow z \in \mathcal{K}_c$ and $Q^k z_n \rightarrow Q^k z$. Hence there are $n_0 \in \mathbb{N}$ and $\sigma > 0$ such that for all $n \geq n_0$,

$$\|I'_n(w)\| \geq \sigma \quad \text{for all } w \in (I_n)_\mu^\nu \setminus C_\delta^n,$$

where $C_\delta^n = C_\delta \cap X_n$. By a standard deformation argument, we can then construct a sequence of odd homeomorphisms $\eta_n : X_n \rightarrow X_n$ such that

$$\eta_n((I_n)_\mu \setminus C_\delta^n) \subset (I_n)_\nu$$

(cf. [12]). For n_0 sufficiently large, we can suppose that

$$\mu < c_n^\ell \leq c_n^{\ell-1} \leq \dots \leq c_n^1 < \nu \quad \text{for all } n \geq n_0.$$

Let $G \in \Sigma_n^\ell$ be such that $\inf I(G) > (\mu + c_n^\ell)/2$. One then has

$$\eta_n(G \setminus C_\delta^n) \subset (I_n)_\nu$$

and

$$\begin{aligned} \gamma(\eta_n(G \setminus C_\delta^n)) &= \gamma(G \setminus C_\delta^n) \geq \gamma(G) - \gamma(C_\delta^n) \\ &\geq n + \ell - \gamma(C_\delta^n) \geq n + \ell - 1. \end{aligned}$$

Thus $\eta_n(G \setminus C_\delta^n) \in \Sigma_n^{\ell-1}$ and $\nu \leq \inf I(\eta_n(G \setminus C_\delta^n)) \leq c_n^{\ell-1}$. One finally comes to $\nu \leq c_n^{\ell-1} < \nu$, which is a contradiction. \square

From now on we turn to the system (E). We denote by $|\cdot|_t$ the usual $L^t(\Omega)$ norm for all $t \in [1, \infty]$. For $q > 1$ let $V_q = H_0^1(\Omega)$ if $q \leq 2^*$ and $V_q = H_0^1(\Omega) \cap L^q(\Omega)$, the Banach space equipped with the norm $\|v\|_{V_q} = (|\nabla v|_2^2 + |v|_q^2)^{1/2}$, if $q > 2^*$. Let E_q be the product space $H_0^1(\Omega) \times V_q$ with elements denoted by $z = (u, v)$. We denote the norm in E_q by $\|z\|_q = (|\nabla u|_2^2 + \|v\|_{V_q}^2)^{1/2}$. E_q has the direct sum decomposition

$$E_q = E_q^- \oplus E^+, \quad z = z^- + z^+$$

where

$$E_q^- = \{0\} \times V_q \quad \text{and} \quad E^+ = H_0^1(\Omega) \times \{0\}.$$

For convenience, we will write $z^+ = u$ and $z^- = v$. Recall that by $(\lambda_n)_{n \in \mathbb{N}}$ we denote the sequence of eigenvalues of $(-\Delta, H_0^1(\Omega))$. Let e_n , $|e_n|_2 = 1$, be the eigenfunction corresponding to λ_n for each $n \in \mathbb{N}$. Clearly, $e_n^+ := (e_n, 0)$, $n \in \mathbb{N}$, is a basis for E^+ , and $e_n^- = (0, e_n)$, $n \in \mathbb{N}$, is a basis for E_q^- .

Suppose that the assumption (H_0) holds. Then

$$(2.3) \quad H(x, z) \leq c(1 + |u|^{2^*} + |v|^q) \quad \text{for all } (x, z).$$

So the functional

$$(2.4) \quad I(z) := \frac{1}{2} \int_{\Omega} (|\nabla u|^2 - |\nabla v|^2) - \int_{\Omega} H(x, z)$$

is well defined in E_q . Moreover, $I \in C^1(E_q, \mathbb{R})$, and the critical points of I are the solutions of (E).

Lemma 2.1. *If (H_0) holds, then I' is weakly sequentially continuous, that is, $I'(z_n) \rightharpoonup I'(z)$ provided $z_n \rightharpoonup z$.*

Proof. If $q < 2^*$ this statement is well known. Assume now that $q \geq 2^*$. Let $z_n \rightharpoonup z$ in E_q . Clearly, for all $w = (\varphi, \psi) \in E_q$, we have

$$\int_{\Omega} (\nabla u_n \nabla \varphi - \nabla v_n \nabla \psi) \rightarrow \int_{\Omega} (\nabla u \nabla \varphi - \nabla v \nabla \psi).$$

So it remains to show that

$$(2.5) \quad \int_{\Omega} H_u(x, z_n) \varphi \rightarrow \int_{\Omega} H_u(x, z) \varphi \quad \text{for all } \varphi \in H_0^1(\Omega)$$

and

$$(2.6) \quad \int_{\Omega} H_v(x, z_n) \psi \rightarrow \int_{\Omega} H_v(x, z) \psi \quad \text{for all } \psi \in V_q.$$

By the Sobolev embedding theorem and using interpolation, we obtain that $u_n \rightarrow u$ in L^t for $t \in [1, 2^*)$ and $v_n \rightarrow v$ in L^t for $t \in [1, q)$. Noting that $|H_u(x, u, v)| \leq \gamma_0(1 + |u|^{p-1} + |v|^{q-1})$ with $2_*(\tau - 1) < q$, (2.5) follows easily since $u_n \rightarrow u$ in L^p , $v_n \rightarrow v$ in $L^{2_*(\tau-1)}$ and $\varphi \in H_0^1(\Omega) \subset L^{2^*}$. Next we see that (2.6) is clearly true when $\psi \in L^\infty$. In general, for a $\psi \in V_q$ we proceed as follows. Let $\tilde{\psi}_m \in L^\infty$ with $\tilde{\psi}_m \rightarrow \psi$ in L^q as $m \rightarrow \infty$. So

$$\left| \int_{\Omega} (H_v(x, z_n) - H_v(x, z)) \psi \right| = \left| \int_{\Omega} (H_v(x, z_n) - H_v(x, z)) (\tilde{\psi}_m + (\psi - \tilde{\psi}_m)) \right|,$$

and using (H_0) we see that this expression is less than the following sum:

$$\begin{aligned} & \left| \int_{\Omega} (H_v(x, z_n) - H_v(x, z)) \tilde{\psi}_m \right| \\ & + c_1 \left(|\psi - \tilde{\psi}_m|_1 + |u_n|_p^{p-1} |\tilde{\psi}_m - \psi|_p + |v_n|_q^{q-1} |\tilde{\psi}_m - \psi|_q \right), \end{aligned}$$

which by its turn is estimated by

$$\left| \int_{\Omega} (H_v(x, z_n) - H_v(x, z)) \tilde{\psi}_m \right| + c_2 \left(|\tilde{\psi}_m - \psi|_p + |\tilde{\psi}_m - \psi|_q \right),$$

since (z_n) is bounded in E_q and L^∞ is dense in L^q . So (2.6) is proved, and it follows that

$$I'(z_n)w \rightarrow I'(z)w \quad \text{for all } w \in E_q.$$

□

3. THE CASE $p > 2$

Throughout this section let (H_0) be satisfied with $p > 2$, and assume that (H_1) and (H_2) hold. Observe that, by (H_2) , there exists $R > 0$ such that $H(x, z) > 0$ whenever $|z| \geq R$. This, jointly with (H_1) , implies

$$(3.1) \quad H(x, z) \geq c_1(|u|^\mu + |v|^\nu) - c_2 \quad \text{for all } (x, z)$$

(see [10]). This, together with (2.3) and (H_2) , shows that

$$(3.2) \quad \nu \leq q \quad \text{and} \quad \beta \leq q.$$

Moreover, by virtue of (3.1) and (H_2) , we may assume, without loss of generality, that (since $\mu > 2$)

$$(3.3) \quad \alpha > 2.$$

Now we set $E^1 = E_q^-$, $E^2 = E^+$ and $e_n^1 = e_n^-$, $e_n^2 = e_n^+$ for all $n \in \mathbb{N}$. So $E_q = E_1 \oplus E_2$. Consider the functional defined by (2.4), which has the properties stated in Section 2.

Lemma 3.1. *Any $(PS)_c^*$ sequence is bounded.*

Proof. Let $z_n \in X_n$ be such that

$$I(z_n) \rightarrow c \quad \text{and} \quad I'_n(z_n) \rightarrow 0.$$

Case 1: $q \leq 2$. In this case $E_q = (H_0^1(\Omega))^2$. By (H_1) , for $w_n := (\frac{1}{\mu}u_n, \frac{1}{\nu}v_n)$, we have

$$(3.4) \quad \begin{aligned} & I(z_n) - I'_n(z_n)w_n \\ &= \left(\frac{1}{2} - \frac{1}{\mu}\right)|\nabla u_n|_2^2 + \left(\frac{1}{\nu} - \frac{1}{2}\right)|\nabla v_n|_2^2 \\ & \quad + \int_{\Omega} \left(\frac{1}{\mu}H_u(x, z_n)u_n + \frac{1}{\nu}H_v(x, z_n)v_n - H(x, z_n) \right) - c_1 \\ & \geq \left(\frac{1}{2} - \frac{1}{\mu}\right)|\nabla u_n|_2^2 + \left(\frac{1}{\nu} - \frac{1}{2}\right)|\nabla v_n|_2^2 - c_2. \end{aligned}$$

If $q < 2$, then (3.2) shows that $\nu < 2$, and so $\|z_n\|_q^2 \leq c_3(1 + \|z_n\|_q)$, which implies that (z_n) is bounded in E_q . Assume $q = 2$. Invoking (3.2), we get $\nu \leq 2$, and so $|\nabla u_n|_2^2 \leq c(1 + \|z_n\|_q)$ by (3.4). Since $H(x, z) > 0$ for all $|z|$ large, and

$$\frac{1}{2}|\nabla v_n|_2^2 + \int_{\Omega} H(x, z_n) = -I(z_n) + \frac{1}{2}|\nabla u_n|_2^2 \leq c(1 + \|z_n\|_q),$$

one sees that $\|z_n\|_q^2 \leq c(1 + \|z_n\|_q)$. Hence, (z_n) is bounded.

Case 2: $q > 2$. Note that in this case $\nu = \mu > 2$ in (H_1) . So

$$(3.5) \quad \begin{aligned} I(z_n) - \frac{1}{2}I'_n(z_n)z_n &= \int_{\Omega} \left(\frac{1}{2}H_z(x, z_n)z_n - H(x, z_n) \right) \\ &\geq \left(\frac{\mu}{2} - 1 \right) \int_{\Omega} H(x, z_n) - c, \end{aligned}$$

which, together with (H_2) , yields

$$(3.6) \quad |u_n|_{\alpha}^{\alpha} + |v_n|_{\beta}^{\beta} \leq c(1 + \|z_n\|_q).$$

Using (H_0) , we get

$$(3.7) \quad \begin{aligned} |\nabla u_n|_2^2 &= I'_n(z_n)(u_n, 0) + \int_{\Omega} H_u(x, z_n) u_n \\ &\leq c_1 \|z_n\|_q + c_2 \int_{\Omega} (|u_n|^p + |v_n|^{\tau-1} |u_n|). \end{aligned}$$

Next we estimate the integrals in the right side of (3.7). Since $2_*(p-1) \leq \alpha \leq p$, we have that $\theta := \alpha/(1+\alpha-p) \leq 2^*$. Using the Hölder inequality, the Sobolev embedding theorem and (3.6), we obtain

$$\int_{\Omega} |u_n|^p \leq |u_n|_{\alpha}^{p-1} |u_n|_{\theta} \leq c_1 + c_2 \|z_n\|_q^{1+(p-1)/\alpha}.$$

Similarly, since $\tau-1 < \beta/2_*$, we have $1 < \omega := \beta/(1+\beta-\tau) < 2^*$, and hence

$$\int_{\Omega} |v_n|^{\tau-1} |u_n| \leq |v_n|_{\beta}^{\tau-1} |u_n|_{\omega} \leq c_1 + c_2 \|z_n\|_q^{1+(\tau-1)/\beta}.$$

Therefore, using the estimate in (3.7), we obtain

$$(3.8) \quad |\nabla u_n|_2^2 \leq c(1 + \|z_n\|_q^{1+(p-1)/\alpha} + \|z_n\|_q^{1+(\tau-1)/\beta}).$$

Since

$$|\nabla v_n|_2^2 = -I'_n(z_n)(0, v_n) - \int_{\Omega} H_z(x, z_n) z_n + \int_{\Omega} H_u(x, z_n) u_n,$$

and using (3.5) and the above arguments, we obtain

$$(3.9) \quad |\nabla v_n|_2^2 \leq c(1 + \|z_n\|_q^{1+(p-1)/\alpha} + \|z_n\|_q^{1+(\tau-1)/\beta}).$$

Recall that, in view of our assumptions, $(p-1)/\alpha \leq 1/2_*$, $(\tau-1)/\beta < 1/2_*$, and $\beta = q$ if $q > 2^*$. Hence, it follows from (3.6) and (3.8)-(3.9) that (z_n) is bounded in E_q . \square

Lemma 3.2. *Let $z_n \in X_n$ be a $(PS)_c^*$ sequence. If $q \leq 2^*$, then (z_n) contains a convergent subsequence. If $q > 2^*$, then there is a $z \in E_q$ such that, along a subsequence, $z_n \rightharpoonup z$ and $I'(z) = 0$ and $I(z) \geq c$.*

Proof. By Lemma 3.1, (z_n) is bounded. We can assume that $z_n \rightharpoonup z$ in E_q , $z_n \rightarrow z$ in $(L^s(\Omega))^2$ for all $1 \leq s < 2^*$, and $z_n(x) \rightarrow z(x)$ a.e. on Ω . It follows from the weak sequential continuity of I' (see Lemma 2.1) that $I'(z) = 0$. Since $I'_n(z_n) \rightarrow 0$, we obtain

$$\begin{aligned} (\nabla u_n, \nabla u_n - \nabla u)_{L^2} &= I'_n(z_n)(u_n - u, 0) + \int_{\Omega} H_u(x, z_n)(u_n - u) \\ &= o(1) + \int_{\Omega} H_u(x, z_n)(u_n - u). \end{aligned}$$

Using (H_0) and the Hölder inequality, we obtain the estimate

$$\begin{aligned} &\left| \int_{\Omega} H_u(x, z_n)(u_n - u) \right| \\ &\leq c \left(|u_n - u|_1 + |u_n|_p^{p-1} |u_n - u|_p + |v_n|_{\beta}^{\tau-1} |u_n - u|_{\omega} \right) = o(1), \end{aligned}$$

where ω is as in the proof of Lemma 3.1. Hence $|\nabla u_n|_2^2 \rightarrow |\nabla u|_2^2$, which implies $u_n \rightarrow u$ in $H_0^1(\Omega)$. Let $P_n : E_q \rightarrow X_n$ denote the projection. Observe that $P_n z \rightarrow z$

in E_q for all $z \in E_q$. Moreover, using again (H_0) and the Hölder inequality, we estimate

$$\begin{aligned} & \left| \int_{\Omega} H_v(x, z_n)(v - P_nv) \right| \\ & \leq c(|v - P_nv|_1 + |u_n|_p^{p-1}|v - P_nv|_p + |v_n|_q^{q-1}|v - P_nv|_q) \rightarrow 0. \end{aligned}$$

On the other hand,

$$\begin{aligned} (\nabla v_n, \nabla v - \nabla v_n)_{L^2} &= o(1) + I'_n(z_n)(0, v_n - P_nv) + \int_{\Omega} H_v(x, z_n)(v_n - Pnv) \\ &= o(1) + \int_{\Omega} H_v(x, z_n)(v_n - v) \\ &= o(1) + \int_{\Omega} H_z(x, z_n)(z_n - z) - \int_{\Omega} H_u(x, z_n)(u_n - u) \\ &= o(1) + \int_{\Omega} H_z(x, z_n)z_n - \int_{\Omega} H_z(x, z_n)z. \end{aligned}$$

Lebesgue's theorem and the weak sequential continuity of $H_z(x, \cdot)$ (see the proof of Lemma 2.1) yield

$$|\nabla v|_2^2 - \limsup_{n \rightarrow \infty} |\nabla v_n|_2^2 = \liminf_{n \rightarrow \infty} \left(\int_{\Omega} H_z(x, z_n)z_n - \int_{\Omega} H_z(x, z_n)z \right) \geq 0,$$

i.e., $|\nabla v|_2^2 \geq \limsup_{n \rightarrow \infty} |\nabla v_n|_2^2$. This, together with the weak lower semicontinuity of norms, implies $|\nabla v_n|_2 \rightarrow |\nabla v|_2$. So $v_n \rightarrow v$ in $H_0^1(\Omega)$.

Therefore, if $q \leq 2^*$, we obtain that, along a subsequence, $z_n \rightarrow z$ in E_q and consequently $I(z) = c$. Next assume that $q > 2^*$. Observe that

$$\begin{aligned} I(z) - I(z_n) &= \frac{1}{2}(|\nabla u|_2^2 - |\nabla u_n|_2^2) - \frac{1}{2}(|\nabla v|_2^2 - |\nabla v_n|_2^2) \\ &\quad + \int_{\Omega} H(x, z_n) - \int_{\Omega} H(x, z); \end{aligned}$$

hence,

$$I(z) - c = o(1) + \int_{\Omega} H(x, z_n) - \int_{\Omega} H(x, z).$$

Lebesgue's theorem then yields

$$I(z) - c = \liminf_{n \rightarrow \infty} \int_{\Omega} H(x, z_n) - \int_{\Omega} H(x, z) \geq 0,$$

that is, $I(z) \geq c$. □

Lemma 3.3. *If (H_3) also holds, there are $r, \rho > 0$ such that $\inf I(\partial B_r E^+) \geq \rho$.*

Proof. By (H_0) and (H_3) , for any $\varepsilon > 0$, there is $c_\varepsilon > 0$ such that

$$H(x, u, 0) \leq \varepsilon|u|^2 + c_\varepsilon|u|^{2^*}.$$

Hence,

$$I(u) \geq \frac{1}{2}|\nabla u|_2^2 - \varepsilon|u|_2^2 - c_\varepsilon|u|_{2^*}^{2^*},$$

and the conclusion follows easily. □

Let $e \in E^+$ with $|\nabla e|_2^2 = 1$, and set

$$Q = \{(se, v) : 0 \leq s \leq r_1, \|v\|_q \leq r_2\}.$$

Lemma 3.4. *If (H_3) also holds, there are $r_1, r_2 > 0$, with $r_1 > r$, such that $I(z) \leq 0$ for all $z \in \partial Q$.*

Proof. By (H_3) , $I(z) \leq 0$ for all $z \in E_q^-$. By (H_2) ,

$$I((se, v)) \leq \frac{s^2}{2} - \frac{1}{2} |\nabla v|_2^2 - c_1 \int_{\Omega} (|se|^{\alpha} + |v|^{\beta}) + c_2.$$

The conclusion follows since $\alpha > 2$. \square

We are now in a position to prove Theorem 1.1.

Proof of Theorem 1.1. Lemmas 3.3 and 3.4 say that I has the linking geometry. Let $Q_n := Q \cap X_n$, and define

$$c_n := \inf_{\gamma \in \Gamma_n} \max I(\gamma(Q_n)),$$

where $\Gamma_n := \{\gamma \in \mathcal{C}(Q_n, X_n) : \gamma|_{\partial Q_n} = id\}$. Then $\rho \leq c_n \leq \kappa := \sup I(Q)$. A standard deformation argument shows that there is $z_n \in X_n$ such that $|I(z_n) - c_n| \leq 1/n$ and $\|I'_n(z_n)\| \leq 1/n$. So we obtain a $(PS)_c^*$ sequence (z_n) with $c \in [\rho, \kappa]$. Lemma 3.2 implies $z_n \rightharpoonup z$ with $I'(z) = 0$ and $I(z) \geq c$. The proof is complete. \square

We now consider the multiplicity of solutions using Proposition 2.1.

Lemma 3.5. *I satisfies (I_1) .*

Proof. Using (H_2) , we obtain

$$I(z) \leq \frac{1}{2} |\nabla u|_2^2 - \frac{1}{2} |\nabla v|_2^2 - c_1 \int_{\Omega} (|u|^{\alpha} + |v|^{\beta}) + c_2.$$

Since all norms in $\text{span}\{e_1, \dots, e_m\}$ are equivalent, we obtain

$$I(z) \leq -\left(c_3 |\nabla u|_2^{\alpha-2} - \frac{1}{2}\right) |\nabla u|_2^2 - \left(\frac{1}{2} |\nabla v|_2^2 + c_1 |v|^{\beta}\right) + c_2,$$

for all $z = (u, v) \in X^m \simeq \text{span}\{e_1, \dots, e_m\} \times V_q$. So (I_1) follows easily. \square

Lemma 3.6. *I satisfies (I_2) .*

Proof. Since $(X^m)^{\perp} \subset H_0^1(\Omega)$ and $H_0^1(\Omega)$ embeds compactly in $L^p(\Omega)$, we have that $\eta_m > 0$ and $\eta_m \rightarrow 0$ as $m \rightarrow \infty$, where

$$(3.10) \quad \eta_m := \sup_{u \in (X^m)^{\perp} \setminus \{0\}} \frac{|u|_p}{|\nabla u|_2};$$

see Lemma 3.8 in [14]. For $z = (u, 0) \in (X^m)^{\perp}$, it follows from (H_0) that

$$\begin{aligned} I(z) &= \frac{1}{2} |\nabla u|_2^2 - \int_{\Omega} H(x, u, 0) \geq \frac{1}{2} |\nabla u|_2^2 - c_1 |u|_p^p - c_2 \\ &\geq \frac{1}{2} |\nabla u|_2^2 - c_1 \eta_m^p |\nabla u|_2^p - c_2. \end{aligned}$$

Setting $r_m = (pc_1 \eta_m^p)^{1/(2-p)}$ and $a_m = (p-2)r_m^2/2p - c_2$, we come to the desired conclusion. \square

Proof of Theorem 1.2. Since $H(x, z)$ is even in z , I is even. Lemma 3.2 shows that I satisfies the assumption (I_4) of Proposition 2.1. Lemmas 3.5 and 3.6 show that (I_1) and (I_2) hold. Clearly (I_3) is also true. Therefore by Proposition 2.1, there is a sequence $(z_n) \subset E_q$ satisfying $I'(z_n) = 0$ and $I(z_n) \rightarrow \infty$. The proof is complete. \square

4. THE CASE $p < 2$

Throughout this section we assume that (H_0) is satisfied with $p \in (1, 2)$. We also suppose that $(H_4) - (H_6)$ hold.

Let $E_q = E^1 \oplus E^2$ be as in Section 3. Consider the functional

$$J(z) = -I(z) = \int_{\Omega} H(x, z) + \frac{1}{2} |\nabla v|_2^2 - \frac{1}{2} |\nabla u|_2^2.$$

Lemma 4.1. *Any $(PS)_c^*$ sequence (z_n) has a subsequence converging weakly to a critical point z of J with $J(z) \leq c$, and $z = 0$ only if $z_n \rightarrow 0$ in E_q .*

Proof. The proof is divided into two parts.

Part I. The sequence (z_n) is bounded in E_q . By (H_4) it follows that

$$J(z_n) - J'_n(z_n) \left(\frac{1}{\mu} u_n, \frac{1}{\nu} v_n \right) \geq \left(\frac{1}{2} - \frac{1}{\nu} \right) |\nabla v_n|_2^2 + \left(\frac{1}{\mu} - \frac{1}{2} \right) |\nabla u_n|_2^2 - c.$$

Hence $|\nabla u_n|_2^2 \leq c(1 + \|z_n\|_q)$. If $\nu > 2$, we also get $|\nabla v_n|_2^2 \leq c(1 + \|z_n\|_q)$. If $\nu = 2$, we use (H_5) and the fact that $|\nabla v|_2^2 \geq \lambda_1 |v|_2^2$ in order to obtain

$$\left(\frac{1}{2} - \delta \right) |\nabla v_n|_2^2 \leq \frac{1}{2} |\nabla v_n|_2^2 + \int_{\Omega} H(x, z_n) = J(z_n) + \frac{1}{2} |\nabla u_n|_2^2.$$

Hence, $|\nabla v_n|_2^2 \leq c(1 + \|z_n\|_q)$, and we get

$$|\nabla u_n|_2^2 + |\nabla v_n|_2^2 \leq c(1 + \|z_n\|_q).$$

Thus, if $q \leq 2^*$, then (z_n) is bounded in E_q . Assume next that $q > 2^*$. It follows from (H_6) that

$$(4.1) \quad J'_n(z_n)(0, v_n) \geq c_1 |v_n|_q^q + |\nabla v_n|_2^2 - c_2 (|v_n|_1 + |u_n|_2^2).$$

Thus $|\nabla u_n|_2^2 + |\nabla v_n|_2^2 + |v_n|_q^q \leq c(1 + \|z_n\|_q)$, which implies that (z_n) is bounded in E_q also in the case when $q > 2^*$.

Part II. We can now suppose that $z_n \rightharpoonup z$ in E_q , $z_n \rightarrow z$ in $(L^s(\Omega))^2$ for all $1 \leq s < 2^*$, and $z_n(x) \rightarrow z(x)$ a.e. in $x \in \Omega$. It follows that z is a critical point of J . As in the proof of Lemma 3.2, using (H_0) and

$$J'_n(z_n)(u_n - u, 0) = \int_{\Omega} H_u(x, z_n)(u_n - u) - (\nabla u_n, \nabla(u_n - u))_{L^2},$$

we obtain that

$$\begin{aligned} & |(\nabla u_n, \nabla(u_n - u))_{L^2}| \\ & \leq o(1) + c(|u_n - u|_1 + |u_n|_p^{p-1} |u_n - u|_p + |v_n|_{\beta}^{\tau-1} |u_n - u|_{\omega}) = o(1), \end{aligned}$$

and so $u_n \rightarrow u$ in $H_0^1(\Omega)$. Let $P_n : E_q \rightarrow X_n$ be the projection as in the proof of Lemma 3.2. So we obtain

$$\begin{aligned} (\nabla v_n, \nabla(v - v_n))_{L^2} &= o(1) + (\nabla v_n, \nabla(P_n v - v_n))_{L^2} \\ &= o(1) + \int_{\Omega} H_v(x, z_n)(v_n - P_n v) - J'_n(z_n)(0, v_n - P_n v) \\ &= o(1) + \int_{\Omega} H_v(x, z_n)(v_n - v) + \int_{\Omega} H_v(x, z_n)(v - P_n v). \end{aligned}$$

Using (H_0) , we have

$$\left| \int_{\Omega} H_v(x, z_n)(v - P_n v) \right| \leq c \left(1 + |u_n|_{2^*}^{2^*} + |v_n|_q^{q-1} \right) \|v - P_n v\|_q \rightarrow 0.$$

Consequently,

$$(4.2) \quad (\nabla v_n, \nabla(v - v_n))_{L^2} = \int_{\Omega} H_v(x, z_n)(v_n - v) + o(1).$$

Thus if $q < 2^*$, it follows from (4.2) that $|\nabla v_n|_2 \rightarrow |\nabla v|_2$, which implies $v_n \rightarrow v$, and so $z_n \rightarrow z$. This proves that J satisfies the $(PS)_c^*$ condition in this case, and that $J(z) = c$.

Consider next $q \geq 2^*$. The weak sequential continuity of $H_v(x, \cdot)$ (see the proof of Lemma 2.1) yields $\int_{\Omega} H_v(x, z_n)v \rightarrow \int_{\Omega} H_v(x, z)v$. By (H_6) , $f_n(x) := H_v(x, z_n)v_n + \gamma_6(|v_n| + |u_n|^2) \geq 0$. Using the fact that $|v_n|_1 \rightarrow |v|_1$ and $|u_n|_2 \rightarrow |u|_2$, and applying Fatou's lemma to the sequence (f_n) , we get

$$\liminf_{n \rightarrow \infty} \int_{\Omega} H_v(x, z_n)v_n \geq \int_{\Omega} H_v(x, z)v.$$

Using this estimate in (4.2), we obtain that $|\nabla v|_2^2 \geq \limsup_{n \rightarrow \infty} |\nabla v_n|_2^2$, which implies that $v_n \rightarrow v$ in $H_0^1(\Omega)$. In order to conclude that $J(z) \leq c$, we use the estimate

$$J(z_n) - J(z) = \int_{\Omega} (H(x, z_n) - H(x, z)) + o(1),$$

(H_4) and Fatou's lemma. Finally, if $z = 0$, then $z_n \rightarrow 0$ in $(H_0^1(\Omega))^2$. By (4.1),

$$|v_n|_q^q \leq o(1) + c(|v_n|_1 + |u_n|_2^2) \rightarrow 0,$$

and so $z_n \rightarrow 0$. □

Remark 4.1. In a similar way, using even simpler arguments, one checks that, if (H_0) holds with $p, q \in (1, 2)$, J satisfies the $(PS)_c^*$ condition for all c .

Remark 4.2. Let $\tilde{J}_m = J|_{X^m}$ denote the restriction of J on X^m . As in Lemma 4.1, it is not difficult to check that, if the sequence $(z_m) \subset E_q$, with $z_m \in X^m$, satisfies $J(z_m) \rightarrow c$ and $\tilde{J}'_m(z_m) \rightarrow 0$ as $m \rightarrow \infty$, then it possesses a subsequence converging weakly to a critical point z of J with $J(z) \leq c$, and $z = 0$ only if $z_n \rightarrow 0$ in E_q . We also have, as in Remark 4.1, that, if (H_0) holds with $p, q \in (1, 2)$, then any such sequence has a convergent subsequence.

Lemma 4.2. *There is an $R > 0$ such that $J(z) \leq 0$ for all $z = (u, 0)$ with $\|z\| \geq R$.*

Proof. By (H_0) , we have $H(x, u, 0) \leq c(1 + |u|^p)$. Hence

$$\begin{aligned} J((u, 0)) &= \int_{\Omega} H(x, u, 0) - \frac{1}{2}|\nabla u|_2^2 \leq c_1 + c_2|u|_p^p - \frac{1}{2}|\nabla u|_2^2 \\ &\leq c_1 - \left(\frac{1}{2}|\nabla u|_2^{2-p} - c_3\right)|\nabla u|_2^p, \end{aligned}$$

and the lemma follows, since $p < 2$. □

Lemma 4.3. *For $\varepsilon > 0$ small there is $\rho > 0$ such that $J((\varepsilon e_1, v)) \geq \rho$ for all $v \in V_q$, where e_1 is the eigenfunction corresponding to the first eigenvalue λ_1 of $(-\Delta, H_0^1(\Omega))$.*

Proof. By (H_5) , for $\varepsilon > 0$ small, $H(x, \varepsilon e_1, v) \geq \gamma_4 \varepsilon^\alpha e_1^\alpha - \delta \lambda_1 v^2$; hence,

$$J((\varepsilon e_1, v)) = \int_{\Omega} H(x, \varepsilon e_1, v) + \frac{1}{2}|\nabla v|_2^2 - \frac{1}{2}\lambda_1 \varepsilon^2 \geq (\gamma_4|e_1|_\alpha^\alpha - \frac{1}{2}\lambda_1 \varepsilon^{2-\alpha})\varepsilon^\alpha.$$

The conclusion follows. □

We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. Recall that $X^m \simeq \text{span}\{e_1, \dots, e_m\} \times V_q$, and consider the restrictions \tilde{J}_m as defined in Remark 4.2. Set $D_R = B_R \cap E^2 = B_R \cap (H_0^1(\Omega) \times \{0\})$ and $D_m = D_R \cap X^m$, where $R > 0$ comes from Lemma 4.2. Define

$$c_m := \inf_{\gamma \in \Gamma_m} \max J(\gamma(D_m)),$$

where $\Gamma_m := \{\gamma \in \mathcal{C}(D_m, X^m) : \gamma(z) = z \text{ for all } z \in \partial D_m\}$. It is well known that $\gamma(D_m) \cap W \neq \emptyset$ for all $\gamma \in \Gamma_m$, where $W = \{(\varepsilon e_1, 0)\} \times V_q$ with $\varepsilon > 0$ small. Invoking Lemma 4.3, we fix an $\varepsilon > 0$ so small that there is $\rho > 0$ satisfying $\inf J(W) \geq \rho$. Then we have

$$\rho \leq c_m \leq b := \max J(D_R).$$

The well-known saddle point theorem (cf. [12] or [4], [14]) implies that there is $z_m \in X^m$ satisfying $|J(z_m) - c_m| \leq 1/m$ and $\|\tilde{J}'_m(z_m)\| \leq 1/m$. Now by virtue of Remark 4.2, along a subsequence, $z_m \rightharpoonup z$ with $J'(z) = 0$ and $z \neq 0$, ending the proof. \square

We now turn to the proof of Theorems 1.4 and 1.5.

Lemma 4.4. *If, in addition, $\gamma_3 = 0$ in (H_4) , then J satisfies (I_5) .*

Proof. It follows from (H_5) that

$$\begin{aligned} J(z) &\geq c_1 |u|_\alpha^\alpha + \left(\frac{1}{2} - \delta\right) |\nabla v|_2^2 - \frac{1}{2} |\nabla u|_2^2 \\ (4.3) \quad &\geq \left(c_2 - \frac{1}{2} |\nabla u|^{2-\alpha}\right) |\nabla u|_2^\alpha + \left(\frac{1}{2} - \delta\right) |\nabla v|_2^2. \end{aligned}$$

Since $\alpha < 2$, the result follows in the case when $q \leq 2^*$. Next consider $q > 2^*$. Suppose (I_5) does not hold. Then for any $r > 0$ there is a sequence $z_j \in X^m$ such that $\|z_j\| = r$ and $J(z_j) \rightarrow 0$. It follows from (4.3) with $z = z_j$, and for r small, that $|\nabla u_j|_2 \rightarrow 0$ and $|\nabla v_j|_2 \rightarrow 0$. All this implies that $\int_\Omega H(x, z_j) \rightarrow 0$. From assumption (H_0) and the fact that (u_j) lies in a finite-dimensional subspace, it follows that $\int_\Omega H_u(x, z_j) u_j \rightarrow 0$. Consequently, by (H_4) with $\gamma_3 = 0$, $\int_\Omega H_v(x, z_j) v_j \rightarrow 0$. This, jointly with (H_6) , yields

$$|v_j|_q^q \leq c_1 \int_\Omega H_v(x, z_j) v_j + c_2 (|v_j|_1 + |u_j|_2^2) \rightarrow 0.$$

Hence, $z_j \rightarrow 0$ in E_q , which is a contradiction. \square

Lemma 4.5. *J satisfies (I_6) .*

Proof. By (H_0) , $H(x, u, 0) \leq c(|u| + |u|^p)$, and so, for $u \in (X^{m-1})^\perp$, one has

$$\begin{aligned} J((u, 0)) &\leq c_1 (|u|_p + |u|_p^p) - \frac{1}{2} |\nabla u|_2^2 \\ &\leq \left(c_1 |u|_p - \frac{1}{4} |\nabla u|_2^2\right) + \left(c_1 |u|_p^p - \frac{1}{4} |\nabla u|_2^2\right) \\ &\leq \left(c_1 \eta_m - \frac{1}{4} |\nabla u|_2^2\right) |\nabla u|_2 + \left(c_1 \eta_m^p - \frac{1}{4} |\nabla u|_2^{2-p}\right) |\nabla u|_2^p, \end{aligned}$$

where η_m was defined by (3.10). Let $b_m := (c_1 \eta_m)^2 + (1-p/2) c_1 \eta_m^p (2p c_1 \eta_m^p)^{p/(2-p)}$. Then $0 < b_m \rightarrow 0$ and $J((u, 0)) \leq b_m$ for all $(u, 0) \in (X^{m-1})^\perp$. \square

Proof of Theorem 1.4. Since $H(x, z)$ is even in z , J is even. If $q \leq 2^*$, then J satisfies the $(PS)_c^*$ condition for all c (see the proof of Lemma 4.1). If $q > 2^*$, then, using assumption (H_4) applied to a critical point z , we obtain

$$J(z) = J(z) - J'_n(z)\left(\frac{1}{\mu}u, \frac{1}{\nu}v\right) \geq \left(\frac{1}{2} - \frac{1}{\nu}\right)|\nabla v|_2^2 + \left(\frac{1}{\mu} - \frac{1}{2}\right)|\nabla u|_2^2 \geq 0.$$

This, jointly with Lemma 4.1, shows that (I_7) is satisfied. It follows from Lemmas 4.4 and 4.5 that J satisfies (I_5) and (I_6) . Therefore, the desired conclusion follows. \square

Finally, we prove Theorem 1.5.

Proof of Theorem 1.5. The proof of the existence of one nontrivial solution is similar to that of Theorem 1.3, using Remark 4.2 and Lemmas 4.2 and 4.3. The other conclusion can be obtained along the lines of the proof of Theorem 1.4, using Remark 4.1 and Lemmas 4.4 and 4.5. \square

5. THE CASE $p = 2$

In this section we always assume that (H_0) holds with $p = 2$ and $\tau < 1 + q/2$. We also suppose that (H_7) and (H_8) are satisfied. We will apply Proposition 2.3 in order to prove Theorem 1.6. Thus, set

$$E^2 = \text{span}\{e_1^+, \dots, e_k^+\} \oplus E_q^- \simeq \text{span}\{e_1, \dots, e_k\} \times V_q, \quad E^1 = E_q \ominus E^2,$$

and

$$X^\ell = E^1 \oplus \text{span}\{e_i^+, \dots, e_k^+, e_1^-, \dots, e_j^-\}.$$

One may arrange the bases as $e_n^1 = e_{k+n}^+$ for $n \in \mathbb{N}$, and $e_n^2 = e_{n+i-1}^+$ for $1 \leq n \leq \ell - j$, $e_n^2 = e_{n-\ell+j}^-$ for $\ell - j < n \leq \ell$, $e_n^2 = e_{n-\ell}^+$ for $\ell < n \leq \ell + i - 1$, and $e_n^2 = e_{n-k}^-$ for $n > \ell + i - 1$. Consider the functional I given by (2.4).

Lemma 5.1. *I satisfies (I_8) ; that is, there exist $r, a > 0$ such that $I(z) \geq a$ for all $z \in X^\ell$ with $\|z\|_q = r$.*

Proof. Let $z = (u, v) \in X^\ell$. Since $v \in \text{span}\{e_1, \dots, e_j\}$, we have $v \in L^\infty$. By (H_0) and (H_7) , for any $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that

$$R_0(x, z) \leq \varepsilon|z|^2 + c_\varepsilon(|u|^{2^*} + |v|^q).$$

Thus

$$\begin{aligned} I(z) &= \frac{1}{2} (|\nabla u|_2^2 - a_0|u|_2^2) - \frac{1}{2} (|\nabla v|_2^2 - b_0|v|_2^2) - \int_\Omega R_0(x, z) \\ &\geq \frac{1}{2} \left(1 - \frac{a_0}{\lambda_i}\right) |\nabla u|_2^2 + \frac{1}{2} \left(\frac{-b_0}{\lambda_j} - 1\right) |\nabla v|_2^2 - \varepsilon|z|_2^2 - c_\varepsilon (|u|_2^{2^*} + |v|_q^q). \end{aligned}$$

Now the conclusion follows easily. \square

Lemma 5.2. *I satisfies (I_9) ; that is, $\sup I(E^2) < \infty$.*

Proof. For $z \in E^2$ we have, using (H_8) , that

$$\begin{aligned} I(z) &= \frac{1}{2} (|\nabla u|_2^2 - a_\infty |u|_2^2) - \frac{1}{2} |\nabla v|_2^2 - \int_\Omega R_\infty(x, z) \\ &\leq -\frac{1}{2} \left(\frac{a_\infty}{\lambda_k} - 1 \right) |\nabla u|_2^2 - \frac{1}{2} |\nabla v|_2^2 + \gamma_9 |u|_\sigma^\sigma - \gamma_8 |v|_q^q + \gamma_9 |\Omega| \\ &\leq -\left(\frac{1}{2} \left(\frac{a_\infty}{\lambda_k} - 1 \right) |\nabla u|^{2-\sigma} - c_1 \right) |\nabla u|_2^\sigma - \left(\frac{1}{2} |\nabla v|_2^2 + \gamma_8 |v|_q^q \right) + c_2, \end{aligned}$$

which implies that $I(z) \leq 0$ for all $z \in E^2$ with $\|z\|_q$ large. \square

Lemma 5.3. *Let $c > 0$. Then any $(PS)_c$ sequence is bounded.*

Proof. We decompose $H_0^1(\Omega)$ as

$$H_0^1(\Omega) = U^- \oplus U^+, \quad u = u^- + u^+,$$

where $U^- = \text{span}\{e_1, \dots, e_k\}$ and U^+ is the orthogonal complement of U^- in $H_0^1(\Omega)$.

Let (z_n) be a $(PS)_c^*$ sequence. Using the expression of I'_n :

$$I'_n(z_n)u_n^+ = |\nabla u_n^+|_2^2 - a_\infty |u_n^+|_2^2 - \int_\Omega \partial_u R_\infty(x, z_n)u_n^+,$$

plus (H_8) and the Hölder inequality, we obtain

$$\left(1 - \frac{a_\infty}{\lambda_{k+1}}\right) |\nabla u_n^+|_2^2 \leq c_1 |\nabla u_n^+|_2 + \gamma_7 \left(|u_n^+|_1 + |u_n|_\sigma^{\sigma-1} |u_n^+|_\sigma + |v_n|_q^{\tau-1} |u_n^+|_r \right)$$

where $r = q/(1 + q - \tau)$. By assumption, $1 < r < 2$. It then follows from the Sobolev embedding theorems that

$$\left(1 - \frac{a_\infty}{\lambda_{k+1}}\right) |\nabla u_n^+|_2^2 \leq c_2 \left(1 + |u_n|_\sigma^{\sigma-1} + |v_n|_q^{\tau-1}\right) |\nabla u_n^+|_2.$$

Similarly, we deduce that

$$\left(\frac{a_\infty}{\lambda_k} - 1\right) |\nabla u_n^-|_2^2 \leq c_2 \left(1 + |u_n|_\sigma^{\sigma-1} + |v_n|_q^{\tau-1}\right) |\nabla u_n^-|_2.$$

The two previous inequalities imply the estimate

$$(5.1) \quad |\nabla u_n|_2^2 \leq c_3 \left(1 + |u_n|_\sigma^{2(\sigma-1)} + |v_n|_q^{2(\tau-1)}\right).$$

Using the expression of H given in (H_8) , and recalling that $I(z_n) > 0$ for large n , we obtain

$$(5.2) \quad \frac{1}{2} |\nabla v_n|_2^2 + \int_\Omega R_\infty(x, z_n) = \frac{1}{2} |\nabla u_n|_2^2 - \frac{a_\infty}{2} |u_n|_2^2 - I(z_n) \leq \frac{1}{2} |\nabla u_n|_2^2.$$

Next using (5.2), assumption (H_8) and (5.1), we obtain

$$(5.3) \quad |\nabla v_n|_2^2 + c_4 |v_n|_q^q \leq c_5 \left(1 + |u_n|_\sigma^\sigma + |u_n|_\sigma^{2(\sigma-1)} + |v_n|_q^{2(\tau-1)}\right).$$

The combination of (5.1) and (5.3) implies

$$|\nabla z_n|_2^2 + |v_n|_q^q \leq c_6 \left(1 + |u_n|_\sigma^\sigma + |v_n|_q^{2(\tau-1)}\right).$$

Since $\sigma < 2$ and $2(\tau - 1) < q$, we see that (z_n) is bounded. \square

Lemma 5.4. *I satisfies (I_{10}) .*

Proof. Let (z_n) be a $(PS)_c^*$ sequence with $c > 0$. Using Lemma 5.3, an argument similar to that of Lemma 3.2 shows that along a subsequence $z_n \rightharpoonup z \in \mathcal{K}_c$, we have $u_n \rightarrow u$ in $H_0^1(\Omega)$. Since $E^1 \subset H_0^1(\Omega)$, we have $P^1 z_n \rightarrow P^1 z$. \square

Proof of Theorem 1.6. Since $H(x, z)$ is even in z , I is even. By assumption, $I(0) = 0$. Lemmas 5.1, 5.2 and 5.4 show that I satisfies $(I_8) - (I_{10})$. Now Proposition 2.3 applies, and the proof is complete. \square

ACKNOWLEDGMENTS

De Figueiredo was supported by CNPq-FAPESP-PRONEX. Ding was supported by the Special Funds for Major State Basic Research Projects of China, the funds of CAS/China R9902, 1001800, and the CNPq of Brazil.

REFERENCES

1. T. Bartsch, *Infinitely many solutions of a symmetric Dirichlet problem*, Nonlinear Anal. TMA, **20** (1993), 1205–1216. MR **94g**:35093
2. T. Bartsch and D. G. De Figueiredo, *Infinitely many solutions of nonlinear elliptic systems*, Progress in Nonlinear Differential Equations and Their Applications, vol. 35, Birkhäuser, Basel/Switzerland, 1999, pp. 51–67. MR **2000j**:35072
3. V. Benci and P. Rabinowitz, *Critical point theorems for indefinite functionals*, Invent. Math. **52** (1979), 241–273. MR **80i**:58019
4. K. C. Chang, *Infinite-Dimensional Morse Theory and Multiple Solution Problems*, Birkhäuser, Boston, 1993. MR **94e**:58023
5. D. G. Costa and C. A. Magalhães, *A variational approach to noncooperative elliptic systems*, Nonlinear Anal. TMA **25** (1995), 699–715. MR **96g**:35070
6. D. G. Costa and C. A. Magalhães, *A unified approach to a class of strongly indefinite functionals*, J. Differential Equations **125** (1996), 521–547. MR **96m**:58061
7. D. G. De Figueiredo and C. A. Magalhães, *On nonquadratic Hamiltonian elliptic systems*, Adv. Differential Equations **1** (1996), 881–898. MR **97f**:35049
8. D. G. De Figueiredo and P. L. Felmer, *On superquadratic elliptic systems*, Trans. Amer. Math. Soc. **343** (1994), 97–116. MR **94g**:35072
9. Y. H. Ding, *Infinitely many entire solutions of an elliptic system with symmetry*, Topological Methods in Nonlinear Anal. **9** (1997), 313–323. MR **99a**:35062
10. P. L. Felmer, *Periodic solutions of “superquadratic” Hamiltonian systems*, J. Differential Equations **102** (1993), 188–207. MR **94c**:58160
11. J. Hulshof and R. van der Vorst, *Differential systems with strongly indefinite variational structure*, J. Funct. Anal. **114** (1993), 32–58. MR **94g**:35073
12. P. H. Rabinowitz, *Minimax Methods in Critical Point Theory with Applications to Differential Equations*, C. B. M. S. vol. 65, Amer. Math. Soc., Providence, RI, 1986. MR **87j**:58024
13. E. A. B. Silva, *Nontrivial solutions for noncooperative elliptic systems at resonance*, Electronic J. Differential Equations **6** (2001), 267–283. MR **2001j**:35097
14. M. Willem, *Minimax Theorems*, Progress in Nonlinear Differential Equations and their Applications, vol. 24, Birkhäuser, Boston, 1996. MR **97h**:58037

IMECC-UNICAMP, CAIXA POSTAL 6065, 13083-970 CAMPINAS S.P. BRAZIL
E-mail address: djairo@ime.unicamp.br

INSTITUTE OF MATHEMATICS, AMSS, CHINESE ACADEMY OF SCIENCES, 100080 BEIJING, PEOPLE'S REPUBLIC OF CHINA
E-mail address: dingyh@math03.math.ac.cn

STABILITY OF SMALL AMPLITUDE BOUNDARY LAYERS FOR MIXED HYPERBOLIC-PARABOLIC SYSTEMS

F. ROUSSET

ABSTRACT. We consider an initial boundary value problem for a symmetrizable mixed hyperbolic-parabolic system of conservation laws with a small viscosity ε , $u_t^\varepsilon + F(u^\varepsilon)_x = \varepsilon(B(u^\varepsilon)u_x^\varepsilon)_x$. When the boundary is noncharacteristic for both the viscous and the inviscid system, and the boundary condition dissipative, we show that u^ε converges to a solution of the inviscid system before the formation of shocks if the amplitude of the boundary layer is sufficiently small. This generalizes previous results obtained for B invertible and the linear study of Serre and Zumbrun obtained for a pure Dirichlet's boundary condition.

1. INTRODUCTION

We consider a one-dimensional system of conservation laws with a small parameter ε set in the domain $x > 0$,

$$(1) \quad u_t^\varepsilon + F(u^\varepsilon)_x = \varepsilon(B(u^\varepsilon)u_x^\varepsilon)_x, \quad x > 0, \quad t > 0,$$

where $u^\varepsilon \in \mathbb{R}^n$ and $F : \mathcal{U} \rightarrow \mathbb{R}^n$, $B : \mathcal{U} \rightarrow \mathbb{R}^{n \times n}$. We will assume that F and B are smooth (C^∞). We add to this system an initial condition $u^\varepsilon(0, x) = u_0(x)$ and a boundary condition that we will detail later. We assume that the eigenvalues of B have nonnegative real part and that the rank of B does not depend on u . We will denote it by r , $1 \leq r \leq n$. Note that B is not necessarily invertible. We are interested in the limit of u^ε when ε tends to zero. We expect that u^ε tends to a solution of the inviscid problem:

$$(2) \quad u_t + F(u)_x = 0$$

with some boundary conditions to be determined. At first we make the natural assumptions to ensure the well-posedness of the Cauchy problem for (1) [6]. There exists a change of variable $u \rightarrow v(u)$ with inverse $u = g(v)$ in which the system can be rewritten as

$$(3) \quad g(v)_t + f(v)_x = \varepsilon(b(v)v_x)_x$$

with the following properties:

- **(H1)** $b(v)$ is block diagonal,

$$b(v) = \begin{pmatrix} 0 & 0 \\ 0 & b_1(v) \end{pmatrix},$$

with $b_1(v) \in GL_r(\mathbb{R})$.

- **(H2)** $dg(v)$ is lower block diagonal,

$$dg(v) = \begin{pmatrix} v & 0 \\ \cdot & \tilde{g}(v) \end{pmatrix},$$

with $\tilde{g}(v) \in GL_r(\mathbb{R})$.

By analogy to the terminology in gas dynamics, we shall refer to v as the primitive variable.

Next we assume that (1) is symmetrizable mixed hyperbolic-parabolic:

- **(H3)** there exists a positive definite symmetric $\Sigma(u)$ such that

(1) $\Sigma(u)dF(u)$ is symmetric,

(2) $\Sigma(u)B(u)X \cdot X \geq \alpha|B(u)X|^2, \forall X \in \mathbb{R}^n$,

where $\alpha > 0$, and \cdot stands for the scalar product of \mathbb{R}^n .

We denote by $v = (w, z)$ the corresponding block decomposition of v .

Note that since $df(v) = dF(u)dg(v)$, $b(v) = B(u)dg(v)$, setting $S(v) = dg(v)^t \Sigma(v)$, we get that (H3) is equivalent to

- **(H3')** There exists $S(v)$ such that

(1) $S(v)dg(v)$ is positive definite symmetric,

(2) $S(v)df(v)$ is symmetric,

(3) $S(v)b(v)X \cdot X \geq \alpha|z|^2, \forall X = \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbb{R}^n$.

We point out that (H3)(1) implies that the inviscid system (2) is hyperbolic. Moreover (H3)(2) implies that $dg^t \Sigma dg$ is block diagonal (see [11], Lemma 4.1). Hence thanks to (H2), we get

$$(4) \quad S = \begin{pmatrix} S_w(v) & 0 \\ * & S_z(v) \end{pmatrix}$$

where $S_w(v)$ is positive definite symmetric. Consequently, writing the block decomposition of df as

$$df(v) = \begin{pmatrix} h(v) & * \\ * & * \end{pmatrix},$$

we get from (H3')(2) that $S_w(v)h(v)$ is symmetric. This means that the system obtained from (3) by removing the second equation is symmetric-hyperbolic.

Finally, we also assume that the hyperbolic and parabolic modes do couple:

- **(H4)** The kernel of B does not contain any eigenvector of dF .

The structural hypotheses **(H1-H4)** are verified by many physical equations as those of compressible gas dynamics and magnetohydrodynamics.

Next we make hypotheses to deal with the initial boundary value problem. We focus on the case of a noncharacteristic boundary. We assume that the boundary is noncharacteristic for both the viscous (1) and the inviscid (2) systems:

- **(H5)** $dF(u)$ and $h(v)$ are nonsingular.

Note that an inflow or outflow boundary condition makes the boundary noncharacteristic in most cases for the Euler and Navier-Stokes equations. These boundary conditions have a physical meaning since they appear in problems with aperture, such as in oil recovery. The analysis of an impermeable boundary would be different since in this case the boundary is characteristic.

We denote by q the number of eigenvalues of positive real part of $dF(U)$, and by p the number of eigenvalues of positive real part of $h(v)$. An initial boundary

value problem for (1) needs $p + r$ scalar independent boundary conditions, and an initial boundary value problem for (2) needs q independent scalar boundary conditions. We deal with boundary conditions for (3) that are linear with respect to the primitive variable v . We write the boundary condition for (3) as

$$(5) \quad Lv(t, 0) = \begin{pmatrix} lw \\ z \end{pmatrix} (t, 0) = g,$$

where l is a linear map that has rank p and g is a given constant.

In the following, in order to make energy estimates, we assume that the boundary condition (5) is "dissipative":

- **(H6)** $\exists \beta > 0, \forall v \in \mathcal{V}, \forall X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, such that $Lv = g, LX = 0$, and we have

$$S_w(v)h(v)X_1 \cdot X_1 \leq -\beta|X_1|^2.$$

There are physical boundary conditions in the form (5) that satisfy (H6). The case of the isentropic gas dynamics will be studied below.

Note that thanks to hypotheses (H1-H5), we have $p \leq q \leq p + r$ ([12], Corollary 1). Hence in the case $q < p + r$, there is a loss of boundary condition when ε tends to zero. It is due to a fast change of u^ε in a vicinity of the boundary: the boundary layer. In the noncharacteristic case the size of the boundary layer is ε . When ε tends to zero, the expected behaviour of u^ε is ([2], [1], [12])

$$u^\varepsilon \sim u^{int}(t, x) + U(t, \frac{x}{\varepsilon}),$$

where u^{int} is a solution of (2) with the initial condition $u^{int}(0, x) = u_0(x)$ and some boundary conditions that we have to determine. $U(t, z)$ is a boundary layer; it is a solution of a differential problem where the time is only a parameter:

$$(6) \quad \begin{aligned} B(U)U' &= F(U + u^{int}(t, 0)) - F(u^{int}(t, 0)), \\ U(t, +\infty) &= 0, \\ Lv(U(t, 0) + u^{int}(t, 0)) &= g. \end{aligned}$$

Note that when $r < n$, we have an algebraic differential system. This problem has solutions if and only if $u^{int}(t, 0)$ belongs to the subset \mathcal{C} , where

$$\mathcal{C} = \left\{ U_+ \in \mathbb{R}^n, \exists U, \begin{cases} B(U)U' = F(U + U_+) - F(U_+), \\ U(+\infty) = 0, \\ Lv(U(0) + U_+) = g. \end{cases} \right\}$$

This set \mathcal{C} is called the set of residual boundary conditions. It was studied in the case $r = n$ in [2], [4] and in the general case in [11]. Assuming that $u_0(0)$ satisfies the boundary condition (5), i.e.,

$$(7) \quad Lv(u_0(0)) = 0,$$

we have $u_0(0) \in \mathcal{C}$ (the associated profile of the boundary layer is $U = 0$). Moreover, thanks to (H1-H6) we can use [11], Lemma 4.2 and Theorem 1.1. \mathcal{C} is a smooth submanifold in the vicinity of $u_0(0)$ that has dimension q and that is transverse to the unstable subspace of $dF(u_0(0))$. Consequently, thanks to a theorem of [8], there exists a continuous solution of (2) with the boundary condition $u^{int}(t, 0) \in \mathcal{C}$ defined on $[0, T]$ for some positive small time T . Assuming some higher-order compatibilities between $u_0(0)$ and \mathcal{C} , we can even get a smooth solution u . Using

the same method as in [4], we can show the existence of an approximate solution of (1) in the form

$$(8) \quad u^{app}(t, x) = u^{int}(t, x) + U(t, \frac{x}{\varepsilon}) + \sum_{i=1}^M \varepsilon^i \left(u^{int, i}(t, x) + U^i(t, \frac{x}{\varepsilon}) \right)$$

such that

$$\begin{aligned} Lv(u^{app}(t, 0)) &= 0, & u^{app}(0, x) &= u_0(x), \\ \partial_t u^{app} + \partial_x F(u^{app}) - \varepsilon \partial_x (B(u^{app}) \partial_x u^{app}) &= R^\varepsilon \end{aligned}$$

where

$$\|R^\varepsilon\|_{L^\infty[0, T], L^2} \leq C\varepsilon^M.$$

Our aim is to show that the true solution u^ε is close to the approximate solution if the boundary layer is sufficiently weak. More precisely,

Theorem 1 (Nonlinear stability). *Assuming (H1-H6) and that $u_0 \in H^7(\mathbb{R}_+)$, there exists $\delta > 0$ such that if*

$$(9) \quad \sup_{t \in [0, T]} \left(|\partial_z U(t, \cdot)|_\infty + \int_0^{+\infty} z |\partial_z U(t, z)| dz + \int_0^{+\infty} |\partial_z U(t, z)| dz \right) \leq \delta,$$

then

$$u^\varepsilon - u^{int} \rightarrow 0$$

when $\varepsilon \rightarrow 0$ in $L^\infty([0, T], L^2)$.

To prove this theorem, we actually need to start from a very accurate approximate solution u^{app} . Indeed, we will take $M = 3$ in the expansion (8). The construction of such a high-order expansion requires a lot of regularity on u^{int} (see [4]). This is why we have to assume so much regularity on u_0 . We actually get a more precise estimate:

$$\|(u^\varepsilon - u^{app})(t, \cdot)\|_{L^2}^2 + \varepsilon^2 \|\partial_x(u^\varepsilon - u^{app})(t, \cdot)\|_{L^2}^2 + \varepsilon^5 \|(u^\varepsilon - u^{app})(t, \cdot)\|_{W^{1, \infty}}^2 \leq C\varepsilon^6.$$

Our method can also provide estimates in $L^\infty([0, T], H^s)$ for any s .

The proof of Theorem 1 relies on energy estimates. We use the primitive variable v ; hence we work on the form (3) of the equation. We combine the energy estimate of the totally parabolic case ($r = n$) [2], [4] with an energy estimate of Kawashima's type [6] and a careful study of the boundary values. A key argument of the proof is the following lemma of [13]:

Lemma 2 (S-K [13]). *Assuming (H1-H4), there exists a skew-symmetric $K(u)$ and a positive constant θ such that*

$$\left((K(u)dF(u))^s + \Sigma(u)B(u) \right) X \cdot X \geq \theta |X|^2, \quad \forall X \in \mathbb{R}^n, \forall u \in \mathcal{U}$$

where $(KdF)^s = \frac{1}{2} (KdF + (KdF)^t)$.

Note that setting $k(v) = dg(v)^t K(g(v))$ we can rewrite this result as

$$(10) \quad \left((k(v)df(v))^s + S(v)b(v) \right) X \cdot X \geq \theta |X|^2, \quad \forall X \in \mathbb{R}^n$$

with

$$(11) \quad k(v)dg(v) \text{ skew-symmetric.}$$

In [9], Lemma 2 and the estimates of [6] combined with pointwise Green's functions bounds were already used to prove the nonlinear asymptotic stability of weak time-independent viscous shock profiles for (1). The asymptotic stability of a time-independent profile of the boundary layer together with the stability of other nonlinear waves was studied in [10] for the isentropic gas dynamics rewritten as a p system in Lagrangian coordinates.

Let us give an example of an application of our theorem. Consider the isentropic gas dynamics where $v = (\rho, v)$, ρ being the mass density and v the fluid velocity,

$$g(v) = \begin{pmatrix} \rho \\ \rho v \end{pmatrix}, \quad f(v) = \begin{pmatrix} \rho v \\ \rho v^2 + p(\rho) \end{pmatrix}, \quad b(v) = \begin{pmatrix} 0 & 0 \\ 0 & \nu(\rho) \end{pmatrix}.$$

Here we assume that $\nu > 0$ and that $p' > 0$ (hyperbolicity). The sound speed is $c(\rho) = \sqrt{p'(\rho)}$. (H1-H4) are verified; moreover, the eigenvalues of dF are $v \pm c$ and the eigenvalue of (the 1×1 matrix) h is v .

Let us first consider an outflow boundary condition

$$v(t, 0) = v^-$$

with $v^- < 0$. In this case, $l = 0$, and (H6) becomes

$$v^- |X|^2 \leq -\beta |X|^2;$$

hence, it is satisfied. The compatibility condition (7) becomes $v_0(0) = v_-$. It suffices to impose $v_0(0) + c(\rho_0(0)) \neq 0$ to get (H5).

If we consider an outflow boundary condition

$$v(t, 0) = v_-, \quad \rho(t, 0) = \rho_-,$$

where $v_- > 0$, we have $l = Id$, $\text{Ker } l = \{0\}$, and hence (H6) is true. The compatibility condition becomes $\rho_0(0) = \rho_-$, $v_0(0) = v_-$ and hence we get (H7) if $v_- - c(\rho_-) \neq 0$. Moreover, in the case $v^- - c(\rho_-) > 0$, we have $q = p + r = 2$; hence there is no boundary layer and the hypothesis (9) is always satisfied.

For a more general discussion of the various boundary conditions for the non-isentropic gas dynamics, we refer to [12].

As in the totally parabolic case $r = n$, the smallness assumption (9) in Theorem 1 is linked with the stability of the boundary layer. In [12] an example of a large unstable boundary layer is given. To understand the mechanism of instability in the boundary layer, we set $\theta = \frac{t}{\varepsilon}$, $z = \frac{x}{\varepsilon}$, we fix some time τ in u^{app} and we linearize about the leading term of u^{app} with respect to ε . We get the linear system

$$(12) \quad \partial_\theta u = \mathcal{L}_\tau u,$$

$$(13) \quad Ldv(u^{int}(\tau, 0) + U(\tau, 0))u(\theta, 0) = 0,$$

$$(14) \quad u(0, z) = u_0(z)$$

where

$$\begin{aligned} \mathcal{L}_\tau u = & \left(b(u^{int}(\tau, 0) + U(\tau, z))u' + dB(u^{int}(\tau, 0) + U(\tau, z))uU'(\tau, z) \right. \\ & \left. - dF(u^{int}(\tau, 0) + U(\tau, z))u \right)'. \end{aligned}$$

Here $'$ stands for $\frac{\partial}{\partial z}$. We will say that the profile of the boundary layer $u^{int}(\tau, 0) + U(\tau, z)$ for some fixed τ is linearly stable if the solutions of this system tend to zero

when t tends to $+\infty$. The linear stability is linked with the spectral stability as was shown in [15], [9], [14]. Let us define the domain of \mathcal{L}_τ as

$$\mathcal{D}(\mathcal{L}_\tau)=\left\{u=dg(V(\tau,z))v, v=\begin{pmatrix} w \\ z \end{pmatrix}, w\in H^1(\mathbb{R}_+), z\in H^2(\mathbb{R}_+), Lv(t,0)=0\right\}$$

where $V(\tau,z)$ is defined by $g(V(\tau,z))=u^{int}(\tau,0)+U(\tau,z)$.

In [12], it is shown that the essential spectrum of \mathcal{L}_τ is confined in $\{\operatorname{Re}\lambda < 0\}\cup\{0\}$ thanks to (H1-H5). In the unstable half-plane $\{\operatorname{Re}\lambda \geq 0\}\setminus\{0\}$ the spectrum only consists of eigenvalues. Consequently, a necessary condition for the linear stability of the boundary layer is that the operator \mathcal{L}_τ does not have eigenvalues in the unstable half-plane $\{\operatorname{Re}\lambda \geq 0\}$ (spectral stability). An Evans function machinery was developed in [12] to find sufficient conditions of instability.

In the first part, we show that spectral stability holds for weak boundary layers.

Theorem 3 (Spectral stability). *There exists $\delta > 0$ such that, assuming (H1-H6) and*

$$(15) \qquad |\partial_z U(\tau,\cdot)|_\infty + \int_0^{+\infty} z|\partial_z U(\tau,z)| + \int_0^{+\infty} |\partial_z U(\tau,z)|\,dz \leq \delta,$$

then \mathcal{L}_τ does not have eigenvalues in the unstable half-plane $\{\operatorname{Re}\lambda \geq 0\}$.

The proof also relies on energy estimates. We first give a direct proof of Theorem 3 because it seems more enlightening to present the main ingredients of the proof of Theorem 1 in the simpler linear time-independent setting of Theorem 3. This result is not used in the proof of Theorem 1. The result of Theorem 3 could be deduced from direct energy estimates on the time evolutionary sytem (12), (13), (14). Nevertheless it is interesting to study the spectral stability since we can expect that, as in the totally parabolic case, the sharp assumption of spectral stability implies the nonlinear convergence result [5].

Note that our result of Theorem 3 (obtained by a different method) implies the result of the appendix of [12] where only Dirichlet’s boundary conditions were considered for (1).

In the second part, we give the proof of the full nonlinear stability result of Theorem 1.

2. SPECTRAL STABILITY

In this section, we prove Theorem 3. We study the eigenvalue problem

$$(16) \qquad \lambda u - \mathcal{L}_\tau u = 0,$$
$$(17) \qquad Ldv(u^{int}(\tau,0)+U(\tau,0))u(0)=0.$$

Setting $u^{int}(\tau,0)+U(\tau,z)=g(V)$, $u=dg(V)v$ (we omit the dependence with respect to τ in this section since τ is fixed), we rewrite the problem in the primitive variable. Hence we have to study the equation

$$(18) \qquad \lambda A^0 v + Av' - (bv')' = A'v + (Cv)',$$
$$(19) \qquad Lv(0)=\begin{pmatrix} lw(0) \\ z(0) \end{pmatrix}=0$$

where $v = \begin{pmatrix} w \\ z \end{pmatrix}$, $A^0(z) = dg(V)$, $A(z) = df(V)$, $b(z) = b(V)$, and $Ch = db(V)hV'$. Note that we have the estimates

$$(20) \quad |A^{0'}| + |A'| + |C| + |C'| + |S'| \leq M|V'|$$

for some $M > 0$, where $S(z)$ stands for $S(V)$.

Moreover, note that thanks to (H1),

$$(21) \quad \Pi_w(Cv)' = 0,$$

where $\Pi_w \begin{pmatrix} w \\ z \end{pmatrix} = w$.

Let us assume that there exists a nonzero solution of (18), (19); without loss of generality, we assume that

$$(22) \quad ||v|| = 1.$$

In this section, since we deal with functions that take complex values, we denote by $u \cdot v$ the scalar product of \mathbb{C}^n ,

$$u \cdot v = \sum_{i=1}^n u_i \bar{v}_i,$$

and by $|\cdot|^2$ the associated norm. We then define

$$||v||^2 = \int_0^{+\infty} v \cdot v \, dx, \quad (u, v) = \int_0^{+\infty} u(x) \cdot v(x) \, dx.$$

We split the proof of the theorem into several lemmas. We will collect all the estimates at the end of the section to reach our conclusion.

We first give an energy estimate in the same spirit as in the totally parabolic case [2], [4] or in the pure Dirichlet's boundary condition case [12]:

Lemma 4. *Assume that v is a solution of (18), (19) that satisfies (22). Then, when δ is sufficiently small, we have the estimate*

$$(23) \quad \operatorname{Re} \lambda + \beta ||z'||^2 + \alpha |w(0)|^2 \leq C \int_0^{+\infty} |V'| |w|^2,$$

$$(24) \quad \operatorname{Re} \lambda + \beta ||z'||^2 + \alpha |w(0)|^2 \leq C\delta (|w(0)|^2 + ||v'||^2),$$

$$(25) \quad |\operatorname{Im} \lambda| \leq C(\delta + ||v'||).$$

Note that the first estimate (23) gives

$$(26) \quad \operatorname{Re} \lambda \leq C\delta.$$

Proof. We first use the same energy estimate as in the strictly parabolic case [2], [4] and the full Dirichlet case [12]. We take the Hermitian product of (16) by Sv (in this section, we will denote $S(V)$ by S for the sake of simplicity) and we take the real part, getting

$$\operatorname{Re} \lambda (SA^0 v, v) + \operatorname{Re} (SA v', v) - \operatorname{Re} (S(bv')', v) = \operatorname{Re} (SA' v, v) + \operatorname{Re} (S(Cv)', v).$$

Since SA is symmetric, we get

$$\begin{aligned}\operatorname{Re}(SAv', v) &= -\frac{1}{2}((SA)'v, v) - \frac{1}{2}SAv(0) \cdot v(0) \\ &\geq -\frac{1}{2}((SA)'v, v) + \alpha|w(0)|^2\end{aligned}$$

thanks to (H6). Note that

$$|(SA)'v, v| \leq C \int_0^{+\infty} |V'| |v|^2 dx.$$

Next, integrating by parts, we have

$$\operatorname{Re}(S(bv')', v) = -\operatorname{Re}(Sbv', v') - \operatorname{Re}((S'bv', v) - \operatorname{Re} Sbv'(0) \cdot v(0).$$

Thanks to (H3'), we have

$$\operatorname{Re}(Sbv', v') \geq \beta|z'|^2;$$

moreover, we have

$$\operatorname{Re} Sbv'(0) \cdot v(0) = 0$$

thanks to the structure of the matrix b given by (H1) and (19). Using again (H1) and (4), we have

$$((S'bv', v)) = (S'_z b_1 z', z) \leq C \left(\eta \|z'\|^2 + \frac{1}{\eta} \int_0^{+\infty} |V'| |z|^2 \right)$$

for every $\eta > 0$ by using the Young inequality. Moreover, we have

$$|(A'v, v)| \leq C \int_0^{+\infty} |V'| |v|^2 dz$$

and

$$\begin{aligned}(27) \quad |(S(Cv)', v)| &= |(SCv, v') + (S' Cv, v)| \\ &\leq C \int_0^{+\infty} |V'| |v| |z'| \leq C \left(\eta \|z'\|^2 + \frac{1}{\eta} \int_0^{+\infty} |V'| |v|^2 \right)\end{aligned}$$

thanks to (H1) and (20).

Collecting these various inequalities, we have shown

$$\operatorname{Re} \lambda \|v\|^2 + \beta \|z'\|^2 + \alpha |w(0)|^2 \leq C \eta \|z'\|^2 + C(\eta) \int_0^{+\infty} |V'| |z|^2 + C(\eta) \int_0^{+\infty} |V'| |v|^2.$$

To conclude, we first choose $\eta = \frac{\beta}{2}$, then we use $z(0) = 0$ through the inequality $|z(x)|^2 \leq x \|z'\|^2$ to get

$$(28) \quad \int_0^{+\infty} |V'| |z|^2 \leq C \int_0^{+\infty} x |V'| \|z'\|^2 \leq C \delta \|z'\|^2$$

and finally, we absorb the terms $C \eta \|z'\|^2$ and $C(\eta) \int_0^{+\infty} |V'| |z|^2$ in the viscous part $\beta \|z'\|^2$ if δ is sufficiently small. This proves (23).

To get (24), we use

$$(29) \quad \int_0^{+\infty} |V'| |w|^2 \leq C \left(\int_0^{+\infty} x |V'| \|w'\|^2 + |w(0)|^2 \int_0^{+\infty} |V'| \right) \leq C \delta (|w(0)|^2 + \|w'\|^2).$$

To prove (25), we also take the scalar product of (16) by Sv , we take the imaginary part and we only use

$$\operatorname{Im}(Sbv'', v) = -\operatorname{Im}(Sbw', v') - \operatorname{Im}((Sb)'v', v) \leq C(\|z'\|^2 + \|v'\| \|v\|).$$

We get

$$\operatorname{Im} \lambda \|v\|^2 \leq C(\|z'\|^2 + \|v'\| \|v\| + \delta \|v\|^2).$$

To conclude, it suffices to use (23), which gives, in particular,

$$\|z'\|^2 \leq C\delta \|w\|^2$$

and the normalization assumption (22).

In the case of a pure Dirichlet boundary condition, a weighted energy estimate on the hyperbolic part of the system (that is to say on the first $n - r$ equations) was used in [12] to bound the term

$$\int_0^{+\infty} |V'| |w|^2.$$

This estimate was similar to the one used by Goodman [3] for the stability of viscous shock profiles. This was efficient because of the upwind propagation. In our more general setting, we use an energy estimate of "Kawashima's type" [6], [9].

Lemma 5. *Assume that v is a solution of (18), (19) that satisfies (22). Then for sufficiently small δ , we have*

$$(30) \quad \|v'\|^2 \leq C(\|z'\|^2 + \|z''\|^2 + (\operatorname{Re} \lambda)^2 + \delta |w(0)|^2).$$

Proof. We use the matrix k given by (10). We apply k to (16), we take the scalar product by v' and we take the real part. Using $\operatorname{Re}(kAv', v') = ((kA)^s v', v')$, we get

$$\begin{aligned} \operatorname{Re}(\lambda(kA^0 v, v')) + ((kA)^s v', v') &\leq C(\|z''\| \|z'\| + \delta \|v'\| \|z'\| + \int_0^{+\infty} |V'| |z|^2) \\ &\leq C(\|z''\| \|z'\| + \delta \|v'\|^2 + \delta |w(0)|^2). \end{aligned}$$

Here we have used the estimates (28) and (29).

Using that kA^0 is skew-Hermitian, we have $(kA^0 v, v') \in \mathbb{R}$ since

$$(kA^0 v, v') = kA^0 v(0) \cdot v(0) - ((kA^0)'v, v) - (kA^0 v', v) = -(kA^0 v', v).$$

Consequently, we have

$$|\operatorname{Re}(\lambda(kA^0 v, v'))| = |\operatorname{Re}(\lambda)(kA^0 v, v')| \leq C \operatorname{Re} \lambda \|v\| \|v'\|.$$

Since we have the estimate (10)

$$((kA)^s v', v') \geq \theta \|v'\|^2 - C \|z'\|^2,$$

we get

$$\|v'\|^2 \leq C(\|z'\|^2 + \|z''\| \|v'\| + \operatorname{Re} \lambda \|v\| \|v'\| + \delta \|v'\|^2 + \delta |w(0)|^2),$$

and hence choosing $\eta > 0$ sufficiently small, using the Young inequality and (9), we have

$$(31) \quad \|v'\|^2 \leq C(\eta)(\|z'\|^2 + \frac{(\operatorname{Re} \lambda)^2}{\eta} \|v\|^2 + \frac{1}{\eta} \|z''\|^2 + \delta |w(0)|^2).$$

Consequently (30) is proved.

To end the proof of the theorem we would want to estimate $\|z''\|$ with respect to $\|v'\|$ and $\|v\|$.

Lemma 6. *Assume that v is a solution of (18), (19) that satisfies (22). Then, when δ is sufficiently small, we have*

$$(32) \quad \|z''\|^2 \leq C \left(\delta \|v'\|^2 + \delta |w(0)|^2 + |z''(0)|^2 + |w'(0)|^2 + |z'(0)|^2 \right).$$

Proof. We take the derivative of (16), getting the equation

$$\lambda A^0 v' + A v'' - (b v'')' = O(|V'|)(|v| + |z''| + |v'|) + (C v)'.$$

The proof is very similar to the proof of (23), in that we take the scalar product of the equation by $S v'$ and we do an integration by parts. The “boundary” terms do not vanish since $v'(0)$ does not satisfy the boundary condition (19). We just point out that to bound the term $((S(Cv))'', v')$ we also do an integration by parts as in (27) to get an estimate independent of $\|v''\|$.

2.1. Proof of Theorem 3. We now give the proof of Theorem 3. To conclude, we first have to eliminate $z''(0)$ and $v'(0)$ in (32).

We first express $w'(0)$, thanks to the hyperbolic part of equation (18):

$$\lambda w + A_1 w' + A_2 z' = O(|V'|)(|v| + |z'|),$$

where $A = \begin{pmatrix} A_1 & A_2 \\ A_2 & A_4 \end{pmatrix}$. Note that we make a crucial use of (21).

Since the boundary is noncharacteristic for the viscous system, A_1 is nonsingular; moreover, thanks to (23), (25), we have

$$(33) \quad |\lambda|^2 \leq \delta + \|v'\|^2.$$

We deduce

$$(34) \quad \begin{aligned} |w'(0)|^2 &\leq C \left(\delta |w(0)|^2 + |z'(0)|^2 + |w(0)|^2 \|v'\|^2 \right) \\ &\leq C \left(\delta |w(0)|^2 + |z'(0)|^2 + \delta \|v'\|^2 \right) \end{aligned}$$

since thanks to (23), we have $|w(0)|^2 \leq C\delta$.

The next step is to estimate $|z'(0)|$. We use the classical Sobolev inequality

$$(35) \quad |z'(0)|^2 \leq 2 \|z''\| \|z\| \leq \eta \|z''\|^2 + \frac{1}{\eta} \|z\|^2$$

for every $\eta > 0$. Hence it suffices to estimate $z''(0)$ in (32). We use the parabolic part of the equation

$$\lambda \tilde{g}(V)z + A_3 w' + A_4 z' - b_1 z'' = O(|V'|)(|v| + |v'|) + O(|\lambda||w|).$$

We get, thanks to (23), (25), (34), and (35),

$$(36) \quad |z''(0)|^2 \leq C \left(\delta |w(0)|^2 + \left(\delta + \frac{1}{\eta} \right) \|z'\|^2 + \delta \|v'\|^2 + \eta \|z''\|^2 \right).$$

Next, we choose η such that $C\eta < 1$, and we replace (34), (35), (36) in (32), getting

$$(37) \quad \|z''\|^2 \leq C \left(\delta \|v'\|^2 + \|z'\|^2 + \delta |w(0)|^2 \right).$$

Finally, collecting (23), (30) and (37), we have shown that

$$(\operatorname{Re} \lambda)(1 - C\delta) + (\beta - C\delta) \|z'\|^2 + (\alpha - C\delta) |w(0)|^2 \leq 0.$$

Hence if δ is sufficiently small, this gives if $\operatorname{Re} \lambda \geq 0$, $z = 0$ and $w(0) = 0$. The hyperbolic part of the equation then becomes a first-order ordinary differential equation involving only w :

$$w' = (A_1)^{-1}(-\lambda w + O(|V'|)w)$$

with the boundary condition $w(0) = 0$. Consequently we also get $w = 0$. This ends the proof of Theorem 3.

3. NONLINEAR STABILITY

In this section we prove Theorem 1. We use the form (3) of the system. Setting $u^\varepsilon = g(v^\varepsilon)$ and $u^{app} = g(v^{app})$, we have the two systems

$$\begin{aligned} g(v^\varepsilon)_t + (f(v^\varepsilon))_x &= \varepsilon(b(v^\varepsilon)v_x^\varepsilon)_x, \\ Lv^\varepsilon(t, 0) &= g, \\ v^\varepsilon(0, x) &= v_0(x) \end{aligned}$$

and

$$\begin{aligned} g(v^{app})_t + (f(v^{app}))_x &= \varepsilon(b(v^{app})v_x^{app})_x + R^\varepsilon, \\ Lv^{app}(t, 0) &= g, \\ v^{app}(0, x) &= v_0(x). \end{aligned}$$

Setting $v^\varepsilon = v^{app} + v$ (we omit the dependence of v in ε), we rewrite our problem as

$$(38) \quad A^0 \partial_t v + A \partial_x v - \varepsilon \partial_x (b \partial_x v) = R^\varepsilon + M^0 + M^1 + M^2,$$

where

$$\begin{aligned} A^0 &= dg(v^{app} + v), \\ A &= A^1 = df(v^{app} + v), \\ b &= b(v^{app} + v), \\ M^0 &= \left(dg(v^{app} + v) - dg(v^{app}) \right) \partial_t v^{app}, \\ M^1 &= \left(df(v^{app} + v) - df(v^{app}) \right) \partial_x v^{app}, \\ M^2 &= \varepsilon \partial_x \left((b(v^{app} + v) - b(v^{app})) \partial_x v^{app} \right). \end{aligned}$$

Note that v satisfies the boundary condition

$$(39) \quad Lv(t, 0) = 0$$

and the initial condition

$$(40) \quad v(0, x) = 0.$$

We choose C sufficiently large such that

$$(41) \quad Q^\varepsilon \leq C\varepsilon^N,$$

where

$$\begin{aligned} Q^\varepsilon &= \|R^\varepsilon\|^2 + \varepsilon^2 \|\partial_t R^\varepsilon\|^2 + \varepsilon^2 \|\partial_x R^\varepsilon\|^2 + \varepsilon^4 \|\partial_{tt} R^\varepsilon\|^2 \\ &\quad + \varepsilon^4 \|\partial_{tx} R^\varepsilon\|^2 + |R^\varepsilon(t, 0)|^2 + \varepsilon^2 \|\partial_t R^\varepsilon(t, 0)\|^2 \end{aligned}$$

for some large N which will be chosen later.

To prove Theorem 1, we use the classical continuous induction argument ([3], [7], [5], [9]). Let us define

$$\begin{aligned} E(t) &= \|v(t)\|^2 + \varepsilon^2 \|\partial_t v(t)\|^2 + \varepsilon^2 \|\partial_x v(t)\|^2 + \varepsilon^4 \|\partial_{tt} v\|^2 + \varepsilon^4 \|\partial_{tx} v(t)\|^2 \\ &\quad + \int_0^t \varepsilon \|\partial_x v(s)\|^2 + \varepsilon^3 \|\partial_{tx} v(s)\|^2 + \varepsilon^3 \|\partial_{xx} z\|^2 ds \\ &\quad + \int_0^t \varepsilon^5 \|\partial_{xxt} z(s)\|^2 + \varepsilon^5 \|\partial_{ttt} z(s)\|^2 ds \\ &\quad + \int_0^t |w(s, 0)|^2 + \varepsilon^2 |\partial_t w(s, 0)|^2 + \varepsilon^4 |\partial_{tt} w(t, 0)|^2 ds. \end{aligned}$$

Note that thanks to (39), (40), (41), we have

$$(42) \quad E(0) \leq C\varepsilon^N.$$

Using the classical short-time theory, we define

$$\begin{aligned} T^* = \sup \Big\{ T^\varepsilon \in [0, T], \exists \text{ a solution of (38), (39), (40) on } [0, T^\varepsilon[\\ \text{such that } \forall t \in [0, T^\varepsilon), E(t) \leq \varepsilon^{N_1} \Big\} \end{aligned}$$

where we choose $N_1 < N$. There are two possibilities:

- (1) $T^* = T$,
- (2) $T^* < T$, and $E(T^*) = \varepsilon^{N_1}$.

In the following, we show by an energy estimate that we cannot be in the second case. This will show Theorem 1.

Let us define $a(\frac{x}{\varepsilon})$ as

$$a\left(\frac{x}{\varepsilon}\right) = \sup_{t \in [0, T]} \sup_{2 \geq \alpha \geq 1, 2 \geq \beta \geq 0} \left| \partial_z^\alpha \partial_t^\beta V\left(t, \frac{x}{\varepsilon}\right) \right|.$$

At first we need an elementary lemma about the estimates of the nonlinear quantities that arise in (38).

Lemma 7. $\forall i = 0, 1$,

$$|\partial_t A^i| \leq C(|\partial_t v|_\infty), \quad |\partial_t^2 A^i| \leq C(|\partial_t v|_\infty)(1 + |\partial_{tt} v|),$$

$$|\partial_x A^i| \leq C(|\partial_x v|_\infty) \left(1 + \frac{1}{\varepsilon} a\left(\frac{x}{\varepsilon}\right)\right),$$

$$|\partial_{tx} A^i| \leq C(|\partial_t v|_\infty, |\partial_x v|_\infty) \left(1 + \frac{1}{\varepsilon} a\left(\frac{x}{\varepsilon}\right) + |\partial_{tx} v|\right).$$

Similar estimates hold for A^i replaced by b , $S(v^{app} + v)$ or $k(v^{app} + v)$. Moreover $\forall \alpha \leq 2, \beta \leq 1$,

$$|\partial_t^\alpha \partial_x^\beta M^0| \leq C(|v|_\infty, |\partial_x v|_\infty, |\partial_t v|_\infty) \sum_{\gamma \leq \alpha, \delta \leq \beta} \left(1 + \frac{1}{\varepsilon^{\beta-\delta}} a\left(\frac{x}{\varepsilon}\right)\right) |\partial_t^\alpha \partial_x^\delta v|,$$

$$|\partial_t^\alpha \partial_x^\beta M^1| \leq C(|v|_\infty, |\partial_x v|_\infty, |\partial_t v|_\infty) \sum_{\gamma \leq \alpha, \delta \leq \beta} \left(1 + \frac{1}{\varepsilon^{\beta-\delta+1}} a\left(\frac{x}{\varepsilon}\right)\right) |\partial_t^\alpha \partial_x^\delta v|,$$

$$|\partial_t^\alpha M^2| \leq C(|v|_\infty, |\partial_x v|_\infty, |\partial_t v|_\infty) \sum_{\gamma \leq \alpha, \delta \leq 1} \left(\varepsilon + \frac{1}{\varepsilon^{1-\delta}} a\left(\frac{x}{\varepsilon}\right)\right) |\partial_t^\alpha \partial_x^\delta v|.$$

Remark 8. We point out that $\Pi_w \partial_t^\alpha \partial_x^\beta M^2 = 0$ where $\Pi_w \begin{pmatrix} w \\ z \end{pmatrix} = w$.

We also point out that, actually, we will not use the case $\alpha = 2, \beta = 1$ in Lemma 7.

We now come to the proof of our main theorem 1. In the proof, C stands for a number that is independent of ε but may depend on T .

Since by classical Sobolev embeddings, we have

$$|v|_\infty^2 \leq C \sup_{t \in [0, T^*]} \|v(t)\|_{H^1}^2, \quad |\partial_t v|_\infty^2 \leq C \sup_{t \in [0, T^*]} \|\partial_t v(t)\|_{H^1}^2,$$

$$|\partial_x z|_\infty^2 \leq 2 \int_0^{T^*} |\partial_{tx} z(s)|_\infty |\partial_x z(s)|_\infty \leq C \left(\int_0^{T^*} \|\partial_{tx} z(s)\|_{H^1}^2 + \|\partial_x z(s)\|_{H^1}^2 ds \right)$$

and since by using the hyperbolic part of equation (38) (that is to say, the w component), the noncharacteristic assumption and Remark 8, we have

$$|\partial_x w|_\infty \leq C(|\partial_t w|_\infty + \frac{1}{\varepsilon}|v|_\infty + |\partial_x z|_\infty),$$

we get the estimate

$$(43) \quad |\partial_t v|_\infty^2 + |\partial_x v|_\infty^2 + |v|_\infty^2 \leq \frac{C}{\varepsilon^5} E(T^*) \leq C \varepsilon^{N_1-5}.$$

Consequently, we choose $N \geq 6$ and $5 < N_1 < N$. This allows us to use Lemma 7. Moreover, thanks to (43), we obtain by continuity from (H6) that

$$(44) \quad S_w(v^{app} + v)h(v^{app} + v)X_1 \cdot X_1 \leq -\beta|X_1|^2, \quad \forall X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad LX = 0$$

since v^{app} satisfies $Lv^{app} = g$.

Note that our smallness assumption (9) and (6) imply that

$$(45) \quad \sup_{z \in \mathbb{R}_+} a(z) + \int_0^{+\infty} a(z) dz + \int_0^{+\infty} za(z) dz \leq C\delta.$$

As for the spectral stability, there are four steps in the proof. We first make the energy estimate of the totally parabolic case; next we make an estimate of Kawashima's type and an estimate on the space derivative of equation (38). The final step is to estimate the boundary values. For this, we replace (25) in the time evolutionary setting by an energy estimate on the time derivative of the equation.

At first, let us make the same energy estimate as for the totally parabolic case [2], [4]. Using Lemma 4 and

$$(46) \quad \frac{1}{\varepsilon} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right) |v|^2 \leq \delta \varepsilon \|\partial_x v\|^2 + \delta \varepsilon |w(t, 0)|^2,$$

$$(47) \quad \begin{aligned} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right) |v| |\partial_x v| &\leq C \|\partial_x v\| \left(\int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right) |v|^2 \right)^{\frac{1}{2}} \leq C \|\partial_x v\| (\sqrt{\varepsilon} \delta |w(t, 0)| + \varepsilon \delta \|\partial_x v\|) \\ &\leq C(\varepsilon \delta \|\partial_x v\|^2 + \delta |w(t, 0)|^2), \end{aligned}$$

we easily get, after absorbing the terms $C\delta|w(t, 0)|^2$ by the term $\alpha|w(t, 0)|^2$ on the left-hand side,

$$(48) \quad \partial_t(SA^0 v, v) + \alpha|w(t, 0)|^2 + \beta \varepsilon \|\partial_x z\|^2 \leq C(Q^\varepsilon + \|v\|^2 + \varepsilon \delta \|\partial_x v\|^2),$$

where S stands for $S(u^{app} + v)$.

Note that an estimate such as (47) is needed to bound the terms

$$(\varepsilon S\partial_x(b\partial_x v), v), (M_2, v).$$

Next we replace the estimate (25) of the spectral stability by an estimate on the time derivative $\partial_t v$. Since $\partial_t v$ still satisfies the boundary condition $L\partial_t v = 0$, we can perform the same computation as previously on the time derivative of (38). Thanks to Lemma 7, we get

$$\begin{aligned} (49) \quad & \partial_t(SA^0\partial_tv, \partial_tv) + \alpha|\partial_tw(t, 0)|^2 + \beta\varepsilon||\partial_{tx}z||^2 \\ & \leq C\Big(\frac{1}{\varepsilon^2}Q^\varepsilon + ||v||^2 + ||\partial_tv||^2 + ||\partial_xv||^2 + \frac{1}{\varepsilon}\int_0^{+\infty} a(\frac{x}{\varepsilon})|\partial_tv|^2\,dx\Big). \end{aligned}$$

We do not give more details since all the ideas of the computation have been used. We just point out that we have used that

$$\begin{aligned} \frac{1}{\varepsilon}\int_0^{+\infty} a(\frac{x}{\varepsilon})|v||\partial_tv|\,dx & \leq \frac{C}{\varepsilon}||\partial_tv||\left(\int_0^{+\infty} a(\frac{x}{\varepsilon})|v|^2\,dx\right)^{\frac{1}{2}} \\ & \leq \frac{C}{\varepsilon}||\partial_tv||(\delta\sqrt{\varepsilon}|w(t, 0)| + \delta\varepsilon||\partial_xv||) \\ (50) \quad & \leq C\Big(\delta||\partial_tv||^2 + \delta||\partial_xv||^2 + \frac{\delta}{\varepsilon}|w(t, 0)|^2\Big) \end{aligned}$$

and that to bound the term

$$\varepsilon\Big(S\partial_x(\partial_tb\partial_xv), \partial_tv\Big)$$

we perform an integration by parts and use the block assumption (H1) and (4). This term is then dominated by

$$C(||\partial_tz||^2 + \frac{\delta}{\varepsilon\eta}||\partial_xz||^2 + \varepsilon\eta||\partial_{tx}z||^2)$$

for every $\eta > 0$. We absorb the last factor by the term $\beta\varepsilon||\partial_{tx}z||^2$ in the left-hand side by choosing η sufficiently small.

Note that for the moment, we do not use an inequality similar to (46) to bound terms such as

$$\int_0^{+\infty} a(\frac{x}{\varepsilon})|\partial_tv|^2\,dx.$$

We bound this term by expressing ∂_tv , thanks to equation (38) and by using estimates such as (46) and (47). Then we get

$$(51) \quad \frac{1}{\varepsilon}\int_0^{+\infty} a(\frac{x}{\varepsilon})|\partial_tv|^2\,dx \leq C\Big(\frac{\delta}{\varepsilon}||\partial_xv||^2 + \delta\varepsilon||\partial_{xx}z||^2 + \frac{\delta}{\varepsilon^2}|w(t, 0)|^2\Big).$$

Note that the factor $\frac{\delta}{\varepsilon^2}|w(t, 0)|^2$ comes from

$$\frac{1}{\varepsilon}\int_0^{+\infty} a(\frac{x}{\varepsilon})|M_1|^2\,dx \leq \frac{1}{\varepsilon}\int_0^{+\infty} a(\frac{x}{\varepsilon})\left(1 + \frac{\delta}{\varepsilon^2}\right)|v|^2\,dx.$$

The estimate (46) gives

$$\frac{1}{\varepsilon^3}\int_0^{+\infty} a(\frac{x}{\varepsilon})|v|^2\,dx \leq C\Big(\frac{\delta}{\varepsilon^2}|w(t, 0)|^2 + \frac{\delta}{\varepsilon}||\partial_xv||^2\Big).$$

Replacing (51) in (49) yields

$$(52) \quad \partial_t(SA^0\partial_tv, \partial_tv) + \alpha|\partial_tw(t, 0)|^2 + \beta\varepsilon||\partial_{tx}z||^2 \\ \leq C\left(\frac{1}{\varepsilon^2}Q^\varepsilon + ||v||^2 + ||\partial_tv||^2 + (1 + \frac{\delta}{\varepsilon})||\partial_xv||^2 + \delta\varepsilon||\partial_{xx}z||^2 + \frac{\delta}{\varepsilon^2}|w(t, 0)|^2\right).$$

As for the spectral stability, the next step is to use the “Kawashima” estimate [6], [9]. We apply $k(v^{app} + v)$ to (38) and we take the scalar product by ∂_xv :

$$(53) \quad (kA^0\partial_tv, \partial_xv) + ((kA)^s\partial_xv, \partial_xv) - \varepsilon(k\partial_x(b\partial_xv)\partial_xv) \\ = (kR^\varepsilon + kM^0 + kM^1 + kM^2, \partial_xv).$$

We use the crucial estimate (10). This gives

$$((kA)^s\partial_xv, \partial_xv) \geq \theta||\partial_xv||^2 - C||\partial_{xx}z||^2.$$

Next we write for every $\eta > 0$,

$$(54) \quad \varepsilon|(k\partial_x(b\partial_xv), \partial_xv)| = |\varepsilon(k\partial_xb\partial_xv, \partial_xv) + \varepsilon(kb\partial_{xx}v\partial_xv)| \\ \leq C\left((\varepsilon + \delta)||\partial_xv||^2 + \varepsilon||\partial_{xx}z||\partial_xv\right) \\ \leq C\left((\varepsilon + \delta + \eta)||\partial_xv||^2 + \frac{\varepsilon^2}{\eta}||\partial_{xx}z||^2\right),$$

where here we have used the block structure assumption (H1) and the Young inequality. Using Lemma 7, the Young inequality, and (47) we have

$$(55) \quad |(kM^1, \partial_xv)| \leq C \int_0^{+\infty} \left(1 + \frac{1}{\varepsilon}a\left(\frac{x}{\varepsilon}\right)\right)|v||\partial_xv| dx \\ \leq C\left(\frac{1}{\eta}||v||^2 + (\eta + \delta)||\partial_xv||^2 + \frac{\delta}{\varepsilon}|w(t, 0)|^2\right).$$

By the same method, we get similar estimates for (kM^0, ∂_xv) and (kM^2, ∂_xv) . To handle $(kA^0\partial_tv, \partial_xv)$, we write

$$(kA^0\partial_tv, \partial_xv) = \partial_t(kA^0v, \partial_xv) - (\partial_t(kA^0)v, \partial_xv) - (kA^0v, \partial_{tx}v).$$

Performing an integration by parts in the last factor above, we get

$$(kA^0\partial_tv, \partial_xv) = \partial_t(kA^0v, \partial_xv) - (\partial_t(kA^0)v, \partial_xv) + kA^0(t, 0)v(t, 0) \cdot \partial_tv(t, 0) \\ + (\partial_x(kA^0)v, \partial_tv) + (kA^0\partial_xv, \partial_tv)$$

and hence, since kA^0 is skew-symmetric,

$$(kA^0\partial_tv, \partial_xv) = \frac{1}{2}\left(\partial_t(kA^0v, \partial_xv) - (\partial_t(kA^0)v, \partial_xv) \right. \\ \left. + kA^0(t, 0)v(t, 0) \cdot \partial_tv(t, 0) + (\partial_x(kA^0)v, \partial_tv)\right).$$

To bound $(\partial_x(kA^0)v, \partial_tv)$, we use

$$\frac{1}{\varepsilon} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right)|v||\partial_tv| dx \leq C\left(\frac{1}{\varepsilon^2} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right)|v|^2 dx + \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right)|\partial_tv|^2 dx\right) \\ \leq C\left(\delta||\partial_xv||^2 + \delta\varepsilon^2||\partial_{xx}z||^2 + \frac{\delta}{\varepsilon}|w(t, 0)|^2\right),$$

thanks to (46) and (51).

This yields

$$(56) \quad (kA^0 \partial_t v, \partial_x v) = \frac{1}{2} \partial_t (kA^0 v, \partial_x v) \\ + O\left(\frac{1}{\varepsilon} \|v\|^2 + (\varepsilon + \delta) \|\partial_x v\|^2 + \frac{\eta + \delta}{\varepsilon} |w(t, 0)|^2 + \frac{\varepsilon}{\eta} |\partial_t w(t, 0)|^2 + \delta \varepsilon^2 \|\partial_{xx} z\|^2\right)$$

for any $\eta > 0$ by use of the Young inequality. Consequently, collecting (54), (55), (56), choosing η sufficiently small and using our assumption (9) to absorb $(\varepsilon + \delta) \|\partial_x v\|^2$ in the left-hand side if δ is sufficiently small, we get from (53),

$$(57) \quad \partial_t (kA^0 v, \partial_x v) + \theta \|\partial_x v\|^2 \leq C \left(\frac{1}{\varepsilon} Q^\varepsilon + \frac{1}{\varepsilon} \|v\|^2 + \frac{1}{\varepsilon} (\delta + \eta) |w(t, 0)|^2 \right. \\ \left. + \frac{\varepsilon}{\eta} |\partial_t w(t, 0)|^2 + \frac{\varepsilon^2}{\eta} \|\partial_{xx} z\|^2 \right).$$

In conclusion, it remains to estimate $\|\partial_{xx} z\|^2$. We take the derivative of (38) with respect to x and we perform an energy estimate similar to (23), but now the boundary terms in the integration by parts do not vanish. Using Lemma 7 and estimates such as (46), (47), and (50), we get

$$(58) \quad \partial_t (SA^0 \partial_x v, \partial_x v) + \beta \varepsilon \|\partial_{xx} z\|^2 \\ \leq C \left(\frac{Q^\varepsilon}{\varepsilon^2} + \left(1 + \frac{\delta}{\varepsilon}\right) \|v\|^2 + \left(1 + \frac{\delta}{\varepsilon}\right) \|\partial_x v\|^2 + \frac{\delta}{\varepsilon^2} |w(t, 0)|^2 \right. \\ \left. + |\partial_x v(t, 0)|^2 + \varepsilon^2 |\partial_{xx} z(t, 0)|^2 \right).$$

Note that to estimate the term $(\partial_x M^2, \partial_x v)$ we have performed an integration by parts to avoid that terms involving $\|\partial_{xx} v\|$ appear and that to estimate $(S \partial_x A^0 \partial_t v, \partial_x v)$, we write

$$|(S \partial_x A^0 \partial_t v, \partial_x v)| \leq C \left(\left(1 + \frac{\delta}{\varepsilon}\right) \|\partial_x v\|^2 + \|\partial_t v\|^2 + \frac{1}{\varepsilon} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right) |\partial_t v|^2 dx \right)$$

and we use (51).

As for the spectral stability, the next step is to estimate the boundary values $\partial_x v(t, 0)$ and $\partial_{xx} z(t, 0)$. We first write the analogue of (35),

$$(59) \quad |\partial_{xx} z(t, 0)|^2 \leq C \left(\varepsilon \eta \|\partial_{xx} z\|^2 + \frac{1}{\varepsilon \eta} \|\partial_{xz}\|^2 \right)$$

for some η sufficiently small so that the term

$$\varepsilon \eta \|\partial_{xx} z\|^2$$

will be absorbed by the left-hand side of (58).

To estimate $\partial_x w(t, 0)$, we use the hyperbolic part of the equation, the fact that the boundary is noncharacteristic and Lemma 7, getting the estimate

$$(60) \quad |\partial_x w(t, 0)|^2 \leq C \left(|R^\varepsilon(t, 0)|^2 + |\partial_t w(t, 0)|^2 + |\partial_{xz}(t, 0)|^2 + \left(1 + \frac{\delta}{\varepsilon^2}\right) |w(t, 0)|^2 \right) \\ \leq C \left(|R^\varepsilon(t, 0)|^2 + |\partial_t w(t, 0)|^2 + \left(1 + \frac{\delta}{\varepsilon^2}\right) |w(t, 0)|^2 \right. \\ \left. + \varepsilon \eta \|\partial_{xx} z\|^2 + \frac{1}{\varepsilon \eta} \|\partial_{xz}\|^2 \right).$$

To estimate $\partial_{xx}z(t, 0)$, we use the parabolic part of the equation. This yields

$$(61) \quad \varepsilon^2 |\partial_{xx}z(t, 0)|^2 \leq C \left(|R^\varepsilon(t, 0)|^2 + |\partial_t w(t, 0)|^2 + \varepsilon \eta \|\partial_{xx}z\|^2 + \frac{1}{\varepsilon \eta} \|\partial_x z\|^2 + (1 + \frac{\delta}{\varepsilon^2}) |w(t, 0)|^2 \right).$$

Next, putting (59), (60), (61) in (58) gives the bound

$$(62) \quad \partial_t(SA^0 \partial_x v, \partial_x v) + \varepsilon \|\partial_{xx}z\|^2 \leq C \left(\frac{1}{\varepsilon^2} Q^\varepsilon + (1 + \frac{\delta}{\varepsilon}) \|v\|^2 + (1 + \frac{\delta}{\varepsilon}) \|\partial_x v\|^2 + \frac{\delta}{\varepsilon^2} |w(t, 0)|^2 \right).$$

Finally, we consider (48) + $\varepsilon^2(52)$ + $\varepsilon(57)$ + $\Gamma \varepsilon^2(62)$ with $\Gamma > 0$ sufficiently large and independent of ε . Integrating from 0 to t , we get the estimate

$$(63) \quad \begin{aligned} & \|v(t)\|^2 + \varepsilon^2 \|\partial_t v(t)\|^2 + \varepsilon^2 \Gamma \|\partial_x v(t)\|^2 \\ & + \beta \left(\int_0^t \varepsilon \|\partial_x v(s)\|^2 + \varepsilon^3 \|\partial_{tx}z(s)\|^2 + \Gamma \varepsilon^3 \|\partial_{xx}z(s)\|^2 ds \right) \\ & + \alpha(|w(t, 0)|^2 + \varepsilon^2 |\partial_t w(t, 0)|^2) + \varepsilon(kAv, \partial_x v)(t) \\ & \leq C \left(\int_0^t \|v(s)\|^2 + \varepsilon^2 \|\partial_t v(s)\|^2 ds + \varepsilon^N \right) \end{aligned}$$

thanks to (41), (42), and (9). Note that we have also used that SA^0 is positive definite symmetric. Thanks to Young's inequality, we have

$$\varepsilon |(kAv, \partial_x v)(t)| \leq C \left(\eta \|v(t)\|^2 + \frac{\varepsilon^2}{\eta} \|\partial_x v(t)\|^2 \right)$$

for any $\eta > 0$. Hence choosing η sufficiently small such that $C\eta < 1$ and then Γ sufficiently large such that $\Gamma > \frac{C}{\eta}$, we finally get

$$\begin{aligned} & \|v(t)\|^2 + \varepsilon^2 \|\partial_t v(t)\|^2 + \varepsilon^2 \|\partial_x v(t)\|^2 \\ & + \left(\int_0^t \varepsilon \|\partial_x v(s)\|^2 + \varepsilon^3 \|\partial_{tx}z(s)\|^2 + \varepsilon^3 \|\partial_{xx}z(s)\|^2 ds \right) \\ & + \alpha(|w(t, 0)|^2 + \varepsilon^2 |\partial_t w(t, 0)|^2) \leq C \left(\int_0^t \|v(s)\|^2 + \varepsilon^2 \|\partial_t v(s)\|^2 ds + \varepsilon^N \right). \end{aligned}$$

We obtain a higher-order estimate by using the same scheme of proof. We first come back to (49), but we use an estimate similar to (46) to bound

$$\frac{1}{\varepsilon} \int_0^{+\infty} a\left(\frac{x}{\varepsilon}\right) |\partial_t v|^2 dx$$

by

$$\varepsilon \delta \|\partial_{tx}v\|^2 + \delta |\partial_t w(t, 0)|^2.$$

Next using Lemma 7, we perform estimates analogous to (49), (57), and (62) for $\partial_{tt}v$, $\partial_{tx}v$ and $\partial_{txx}z$ respectively.

Finally, we obtain the estimate

$$E(t) \leq C \left(\varepsilon^N + \int_0^t E(s) ds \right).$$

Therefore, by Gronwall's Lemma we have

$$E(t) \leq C \varepsilon^N \quad \forall t \in [0, T^*].$$

Hence $E(T^*) < \varepsilon^{N_1}$ since $N > N_1$ and hence $T^* = T$.

REFERENCES

- [1] M. Gisclon, *Étude des conditions aux limites pour un système strictement hyperbolique, via l'approximation parabolique*, J. Math. Pures Appl. (9) **75** (1996), no. 5, 485–508. MR **97f**:35129
- [2] M. Gisclon and D. Serre, *Étude des conditions aux limites pour un système strictement hyperbolique via l'approximation parabolique*, C. R. Acad. Sci. Paris Sér. I Math. **319** (1994), no. 4, 377–382. MR **95e**:35119
- [3] J. Goodman, *Nonlinear asymptotic stability of viscous shock profiles for conservation laws*, Arch. Rational Mech. Anal. **95** (1986), no. 4, 325–344. MR **88b**:35127
- [4] E. Grenier and O. Guès, *Boundary layers for viscous perturbations of noncharacteristic quasilinear hyperbolic problems*, J. Differential Equations **143** (1998), no. 1, 110–146. MR **98j**:35026
- [5] E. Grenier and F. Rousset, *Stability of one-dimensional boundary layers by using Green's functions*, Comm. Pure Appl. Math. **54** (2001), no. 11, 1343–1385. MR **2003a**:35126
- [6] S. Kawashima, *Systems of a hyperbolic parabolic type with applications to the equations of magnetohydrodynamics*, Ph.D. thesis, Kyoto University (1983).
- [7] G. Kreiss and H.-O. Kreiss, *Stability of systems of viscous conservation laws*, Comm. Pure Appl. Math. **51** (1998), no. 11–12, 1397–1424. MR **2000c**:35156
- [8] T. T. Li and W. C. Yu, *Boundary value problems for quasilinear hyperbolic systems*, Duke University Mathematics Department, Durham, N.C., 1985. MR **88g**:35115
- [9] C. Mascia and K. Zumbrun, *Stability of viscous shock profiles for dissipative symmetric hyperbolic-parabolic systems*, Preprint (2001).
- [10] A. Matsumura and K. Nishihara, *Large-time behaviors of solutions to an inflow problem in the half space for a one-dimensional system compressible viscous gas*, Commun. Math. Physics **222** (2001), 449–474. MR **2002m**:76083
- [11] F. Rousset, *The boundary conditions coming from the real vanishing viscosity method*, Discrete Continuous Dynamical Systems (to appear).
- [12] D. Serre and K. Zumbrun, *Boundary layer stability in real vanishing viscosity limit*, Commun. Math. Phys. **221** (2001), 267–292.
- [13] Y. Shizuta and S. Kawashima, *Systems of equations of hyperbolic-parabolic type with applications to the discrete Boltzmann equation*, Hokkaido Math. J. **14** (1985), no. 2, 249–275. MR **86k**:35107
- [14] K. Zumbrun, *Multidimensional stability of planar viscous shock waves*, Advances in the theory of shock waves, Prog. Nonlinear Differential Equations Appl. **47**, Birkhäuser, Boston, MA, 307–516 (2001). (English). MR **2002k**:35200
- [15] K. Zumbrun and P. Howard, *Pointwise semigroup methods and stability of viscous shock waves*, Indiana Univ. Math. J. **47** (1998), no. 3, 741–871. MR **99m**:35157

ENS LYON, UMPA (UMR 5669 CNRS), 46, ALLÉE D'ITALIE, 69364 LYON CEDEX 07, FRANCE
E-mail address: frousset@umpa.ens-lyon.fr

Current address: Laboratoire Dieudonné, Université de Nice-Sophia Antipolis, Parc Valrose,
 06108 Nice Cedex 02, France

E-mail address: frousset@math.unice.fr

Editorial Information

To be published in the *Transactions*, a paper must be correct, new, nontrivial, and significant. Further, it must be well written and of interest to a substantial number of mathematicians. Piecemeal results, such as an inconclusive step toward an unproved major theorem or a minor variation on a known result, are in general not acceptable for publication.

Papers submitted to the *Transactions* should exceed 10 published journal pages in length. Shorter papers may be submitted to the *Proceedings of the American Mathematical Society*. Published pages are the same size as those generated in the style files provided for $\text{AMS-LAT}_{\text{E}}\text{X}$ or $\text{AMS-T}_{\text{E}}\text{X}$.

As of February 28, 2003, the backlog for this journal was approximately 2 issues. This estimate is the result of dividing the number of manuscripts for this journal in the Providence office that have not yet gone to the printer on the above date by the average number of articles per issue over the previous twelve months, reduced by the number of issues published in four months (the time necessary for editing and composing a typical issue). In an effort to make articles available as quickly as possible, articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

A Consent to Publish and Copyright Agreement is required before a paper will be published in this journal. After a paper is accepted for publication, the Providence office will send a Consent to Publish and Copyright Agreement to all authors of the paper. By submitting a paper to this journal, authors certify that the results have not been submitted to nor are they under consideration for publication by another journal, conference proceedings, or similar publication.

Information for Authors

Initial submission. Two copies of the paper should be sent directly to the appropriate Editor and the author should keep a copy. *If an editor is agreeable*, an electronic manuscript prepared in $\text{T}_{\text{E}}\text{X}$ or $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ may be submitted by pointing to an appropriate URL on a preprint or e-print server.

The first page must consist of a *descriptive title*, followed by an *abstract* that summarizes the article in language suitable for workers in the general field (algebra, analysis, etc.). The *descriptive title* should be short, but informative; useless or vague phrases such as “some remarks about” or “concerning” should be avoided. The *abstract* should be at least one complete sentence, and at most 300 words. Included with the footnotes to the paper should be the 2000 *Mathematics Subject Classification* representing the primary and secondary subjects of the article. The classifications are accessible from www.ams.org/msc/. The list of classifications is also available in print starting with the 1999 annual index of *Mathematical Reviews*. The Mathematics Subject Classification footnote may be followed by a list of *key words and phrases* describing the subject matter of the article and taken from it. Journal abbreviations used in bibliographies are listed in the latest *Mathematical Reviews* annual index. The series abbreviations are also accessible from www.ams.org/publications/. To help in preparing and verifying references, the AMS offers MR Lookup, a Reference Tool for Linking, at www.ams.org/mrlookup/. When the manuscript is submitted, authors should supply the editor with electronic addresses if available. These will be printed after the postal address at the end of each article.

Electronically prepared manuscripts. The AMS encourages electronically prepared manuscripts, with a strong preference for $\text{AMS-LAT}_{\text{E}}\text{X}$. To this end, the

Society has prepared $\text{\AA MS-L\AA T\AA E X}$ author packages for each AMS publication. Author packages include instructions for preparing electronic manuscripts, the *AMS Author Handbook*, samples, and a style file that generates the particular design specifications of that publication series. Articles properly prepared using the $\text{\AA MS-L\AA T\AA E X}$ style file and the `\label` and `\ref` commands automatically enable extensive intra-document linking to the bibliography and other elements of the article for searching electronically on the Web. Because linking must often be added manually to electronically prepared manuscripts in other forms of \AA T\AA E X , using $\text{\AA MS-L\AA T\AA E X}$ also reduces the amount of technical intervention once the files are received by the AMS. This results in fewer errors in processing and saves the author proofreading time. $\text{\AA MS-L\AA T\AA E X}$ papers also move more efficiently through the production stream, helping to minimize publishing costs.

$\text{\AA MS-L\AA T\AA E X}$ is the highly preferred format of \AA T\AA E X , but author packages are also available in \AA MS-T\AA E X . Those authors who make use of these style files from the beginning of the writing process will further reduce their own efforts. Manuscripts prepared electronically in \AA T\AA E X or plain \AA T\AA E X are normally not acceptable due to the high amount of technical time required to insure that the file will run properly through the AMS in-house production system. \AA T\AA E X users will find that $\text{\AA MS-L\AA T\AA E X}$ is the same as \AA T\AA E X with additional commands to simplify the typesetting of mathematics, and users of plain \AA T\AA E X should have the foundation for learning $\text{\AA MS-L\AA T\AA E X}$.

Authors may retrieve an author package from the AMS website starting from www.ams.org/tex/ or via FTP to [ftp.ams.org](ftp://ftp.ams.org) (login as `anonymous`, enter username as password, and type `cd pub/author-info`). The *AMS Author Handbook* and the *Instruction Manual* are available in PDF format following the author packages link from www.ams.org/tex/. The author package can also be obtained free of charge by sending email to pub@ams.org (Internet) or from the Publication Division, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When requesting an author package, please specify $\text{\AA MS-L\AA T\AA E X}$ or \AA MS-T\AA E X , Macintosh or IBM (3.5) format, and the publication in which your paper will appear. Please be sure to include your complete mailing address.

At the time of submission, authors should indicate if the paper has been prepared using $\text{\AA MS-L\AA T\AA E X}$ or \AA MS-T\AA E X and provide the Editor with a paper manuscript that matches the electronic manuscript. The final version of the electronic manuscript should be sent to the Providence office immediately after the paper has been accepted for publication. The author should also send the final version of the paper manuscript to the Editor, who will forward a copy to the Providence office. Editors will require authors to send their electronically prepared manuscripts to the Providence office in a timely fashion. Electronically prepared manuscripts can be sent via email to pub-submit@ams.org (Internet) or on diskette to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When sending a manuscript electronically, please be sure to include a message indicating in which publication the paper has been accepted. No corrections will be accepted electronically. Authors must mark their changes on their proof copies and return them to the Providence office. Complete instructions on how to send files are included in the author package.

Electronic graphics. Comprehensive instructions on preparing graphics are available starting from www.ams.org/jourhtml/authors.html. A few of the major requirements are given here.

Submit files for graphics as EPS (Encapsulated PostScript) files. This includes graphics originated via a graphics ap_____r

other computer-generated images. If this is not possible, TIFF files are acceptable as long as they can be opened in Adobe Photoshop or Illustrator. No matter what method was used to produce the graphic, it is necessary to provide a paper copy to the AMS.

Authors using graphics packages for the creation of electronic art should also avoid the use of any lines thinner than 0.5 points in width. Many graphics packages allow the user to specify a “hairline” for a very thin line. Hairlines often look acceptable when proofed on a typical laser printer. However, when produced on a high-resolution laser imagesetter, hairlines become nearly invisible and will be lost entirely in the final printing process.

Screens should be set to values between 15% and 85%. Screens which fall outside of this range are too light or too dark to print correctly. Variations of screens within a graphic should be no less than 10%.

AMS policy on making changes to articles after posting. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue. To preserve the integrity of electronically published articles, once an article is individually posted to the AMS website but not yet in an issue, changes cannot be made in place in the paper. However, an “Added after posting” section may be added to the paper right before the References when there is a critical error in the content of the paper. The “Added after posting” section gives the author an opportunity to correct this type of critical error before the article is put into an issue for printing and before it is then reposted with the issue. The “Added after posting” section remains a permanent part of the paper. The AMS does not keep author-related information, such as affiliation, current address, and email address, up to date after a paper is initially posted.

Once the article is assigned to an issue, even if the issue has not yet been posted to the AMS website, corrections may be made to the paper by submitting a traditional errata article to the Editor. The errata article will appear in a future print issue and will link back and forth on the web to the original article online.

Secure manuscript tracking on the Web and via email. Authors can track their manuscripts through the AMS journal production process using the personal AMS ID and Article ID printed in the upper right-hand corner of the Consent to Publish form sent to each author who publishes in AMS journals. Access to the tracking system is available from www.ams.org/mstrack/ or via email sent to mstrack-query@ams.org. To access by email, on the subject line of the message simply enter the AMS ID and Article ID. To track more than one manuscript by email, choose one of the Article IDs and enter the AMS ID and the Article ID followed by the word *all* on the subject line. An explanation of each production step is provided on the web through links from the manuscript tracking screen. Questions can be sent to tran-query@ams.org.

T_EX files available. Beginning with the January 1992 issue of the *Bulletin* and the January 1996 issues of *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS*, T_EX files can be downloaded from the AMS website, starting from www.ams.org/journals/. Authors without Web access may request their files at the address given below after the article has been published. For *Bulletin* papers published in 1987 through 1991 and for *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS* papers published in 1987 through 1995, T_EX files are available upon request for authors without Web access by sending email to file-request@ams.org or by contacting the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2204 USA. The request should include the title of the paper, the

name(s) of the author(s), the name of the publication in which the paper has or will appear, and the volume and issue numbers if known. The \TeX file will be sent to the author making the request after the article goes to the printer. If the requestor can receive Internet email, please include the email address to which the file should be sent. Otherwise please indicate a diskette format and postal address to which a disk should be mailed. **Note:** Because \TeX production at the AMS sometimes requires extra fonts and macros that are not yet publicly available, \TeX files cannot be guaranteed to run through the author's version of \TeX without errors. The AMS regrets that it cannot provide support to eliminate such errors in the author's \TeX environment.

Inquiries. Any inquiries concerning a paper that has been accepted for publication that cannot be answered via the manuscript tracking system mentioned above should be sent to tran-query@ams.org or directly to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

Editors

The traditional method of submitting a paper is to send two hard copies to the appropriate editor. Subjects, and the editors associated with them, are listed below.

In principle the Transactions welcomes electronic submissions, and some of the editors, those whose names appear below with an asterisk (*), have indicated that they prefer them. Editors reserve the right to request hard copies after papers have been submitted electronically. Authors are advised to make preliminary inquiries to editors as to whether they are likely to be able to handle submissions in a particular electronic form.

Algebra and algebraic geometry, KAREN E. SMITH, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109 USA; e-mail: kesmith@umich.edu

Algebraic geometry, DAN ABRAMOVICH, Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215 USA; e-mail: abramovic@bu.edu

Algebraic topology and cohomology of groups, STEWART PRIDDY, Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730 USA; e-mail: priddy@math.nwu.edu

* **Combinatorics**, SERGEY FOMIN, Department of Mathematics, East Hall, University of Michigan, Ann Arbor, MI 48109-1109 USA; e-mail: fomin@umich.edu

Complex analysis and geometry, D. H. PHONG, Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027-0029 USA; e-mail: phong@math.columbia.edu

* **Differential geometry and global analysis**, LISA C. JEFFREY, Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3; e-mail: jeffrey@math.toronto.edu

Dynamical systems and ergodic theory, ROBERT F. WILLIAMS, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: bob@math.utexas.edu

* **Geometric analysis**, TOBIAS COLDING, Courant Institute, New York University, 251 Mercer Street, New York, NY 10012 USA; e-mail: colding@cims.nyu.edu

Geometric topology, knot theory, and hyperbolic geometry, ABIGAIL THOMPSON, Department of Mathematics, University of California, Davis, CA 95616-5224 USA; e-mail: thompson@math.ucdavis.edu

Harmonic analysis, ALEXANDER NAGEL, Department of Mathematics, University of Wisconsin, 480 Lincoln Drive, Madison, WI 53706-1313 USA; e-mail: nagel@math.wisc.edu

Harmonic analysis, representation theory, and Lie theory, ROBERT J. STANTON, Department of Mathematics, Ohio State University, 231 West 18th Avenue, Columbus, OH 43210-1174 USA; e-mail: stanton@math.ohio-state.edu

* **Logic**, THEODORE SLAMAN, Department of Mathematics, University of California, Berkeley, CA 94720-3840 USA; e-mail: slaman@math.berkeley.edu

Number theory, HAROLD G. DIAMOND, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801-2917 USA; e-mail: diamond@math.uiuc.edu

* **Ordinary differential equations, partial differential equations, and applied mathematics**, PETER W. BATES, Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027 USA; e-mail: bates@math.msu.edu

* **Partial differential equations**, PATRICIA E. BAUMAN, Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395 USA; e-mail: bauman@math.purdue.edu

* **Probability and statistics**, KRZYSZTOF BURDZY, Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350 USA; e-mail: burdzy@math.washington.edu

* **Real analysis and partial differential equations**, DANIEL TATARU, Department of Mathematics, University of California, Berkeley, CA 94720 USA; e-mail: tataru@math.berkeley.edu

All other communications to the editors should be addressed to the Managing Editor, WILLIAM BECKNER, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: beckner@math.utexas.edu


MEMOIRS OF THE AMERICAN MATHEMATICAL SOCIETY

Memoirs is devoted to research in pure and applied mathematics of the same nature as *Transactions*. An issue consists of one or more separately bound research tracts for which the authors provide reproduction copy. Papers intended for *Memoirs* should normally be at least 80 pages in length. *Memoirs* has the same editorial committee as *Transactions*; so such papers should be addressed to one of the editors listed above.

TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY



TRANSACTIONS
OF THE
AMERICAN MATHEMATICAL SOCIETY

ISSN 1088-6826 (print) / ISSN 1088-6834 (online)

[Recently posted articles / Most recent issue / All issues](#)

About this journal

- Journal services
- Editorial board
- Creating and submitting a manuscript
- Preparing a paper for publication
- Back issues from 1912


Subscription information

- Subscribe from the AMS Bookstore
- Library/Institutional subscription agreement

For authors

- Initial submission
- Author guidelines
- Consent to publish and copyright agreements
- Permissions
- Ways to read (free for accepted papers)
- Book review information
- Information for authors on submitting electronic manuscripts
- AMS LaTeX system

Submission information:
Information can be found on the journal's Initial Submission page.



TRANSACTIONS
OF THE
AMERICAN MATHEMATICAL SOCIETY

ISSN 1088-6826 (print) / ISSN 1088-6834 (online)

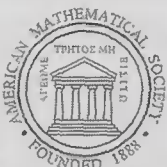
[Recently posted articles / Most recent issue / All issues](#)

Journal overview: This journal is devoted to research articles in all areas of pure and applied mathematics. To be published in the *Transactions*, a paper must be correct, new, and significant. Further, it must be well written and of interest to a substantial number of mathematicians.

The AMS has released enhanced versions of its electronic journals. These upgrades improve usefulness and relevance for both journal subscribers and journal authors.

A 30-day free trial is available to corporations and institutions. A downloadable Free Trial Form is available at: www.ams.org/customers/ejournaltrial.pdf.

Contact AMS Membership and Customer Services, 201 Charles Street, Providence, RI 02904-2294, USA; phone 1-800-321-4267 or 1-401-455-4000 worldwide; fax 1-401-455-4046; email: cust-serv@ams.org. Note: A signed license agreement is required for AMS electronic journal subscriptions. A newly updated and expanded agreement can be found online at <http://www.ams.org/customers/jour-license.html>.



www.ams.org/tran

TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY
CONTENTS

Vol. 355, No. 7 Whole No. 818 July 2003

Borislav Karaivanov, Pencho Petrushev, and Robert C. Sharpley,
Algorithms for nonlinear piecewise polynomial approximation: Theoretical aspects 2585

Jörg Brendle, The almost-disjointness number may have countable cofinality 2633

Alina Carmen Cojocaru, Cyclicity of CM elliptic curves modulo p 2651

Tonghai Yang, Taylor expansion of an Eisenstein series 2663

Eric Freeman, Systems of diagonal Diophantine inequalities 2675

Francisco Javier Gallego and Bangere P. Purnaprajna, On the canonical rings of covers of surfaces of minimal degree 2715

H. H. Brungs and N. I. Dubrovin, A classification and examples of rank one chain domains 2733

Donald W. Barnes, On the spectral sequence constructors of Guichardet and Stefan 2755

Steven Lillywhite, Formality in an equivariant setting 2771

Neil Hindman, Dona Strauss, and Yevhen Zelenyuk, Large rectangular semigroups in Stone-Čech compactifications 2795

Takehiko Yamanouchi, Galois groups of quantum group actions and regularity of fixed-point algebras 2813

Boo Rim Choe, Hyungwoon Koo, and Wayne Smith, Composition operators acting on holomorphic Sobolev spaces 2829

B. Jakubczyk and M. Zhitomirskii, Distributions of corank 1 and their characteristic vector fields 2857

E. Boeckx, When are the tangent sphere bundles of a Riemannian manifold reducible? 2885

Henri Comman, Criteria for large deviations 2905

Seung Jun Chang, Jae Gil Choi, and David Skoug, Integration by parts formulas involving generalized Fourier-Feynman transforms on function space 2925

Michiko Yuri, Thermodynamic formalism for countable to one Markov systems 2949

D. G. De Figueiredo and Y. H. Ding, Strongly indefinite functionals and multiple solutions of elliptic systems 2973

F. Rousset, Stability of small amplitude boundary layers for mixed hyperbolic-parabolic systems 2991



VOLUME 355 NUMBER 8



AUGUST 2010

WHOLE NUMBER

TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

Dan Abramovich

Peter W. Bates

Patricia E. Bauman

William Beckner, Managing Editor

Krzysztof Burdzy

Tobias Colding

Harold G. Diamond

Sergey Fomin

Lisa C. Jeffrey

Alexander Nagel

D. H. Phong

Stewart Priddy

Theodore Slaman

Karen E. Smith

Robert J. Stanton

Daniel Tataru

Abigail Thompson

Robert F. Williams

PROVIDENCE, RHODE ISLAND USA

ISSN 0002-9947

Available electronically

www.ams.org

Transactions of the American Mathematical Society

This journal is devoted entirely to research in pure and applied mathematics.

Submission information. See **Information for Authors** at the end of this issue.

Publisher Item Identifier. The Publisher Item Identifier (PII) appears at the top of the first page of each article published in this journal. This alphanumeric string of characters uniquely identifies each article and can be used for future cataloging, searching, and electronic retrieval.

Postings to the AMS website. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

Subscription information. *Transactions of the American Mathematical Society* is published monthly. Beginning in January 1996 *Transactions* is accessible from www.ams.org/publications/. Subscription prices for Volume 355 (2003) are as follows: for paper delivery, \$1490 list, \$1192 institutional member, \$1341 corporate member; for electronic delivery, \$1341 list, \$1073 institutional member, \$1207 corporate member. Upon request, subscribers to paper delivery of this journal are also entitled to receive electronic delivery. If ordering the paper version, add \$39 for surface delivery outside the United States and India; \$50 to India. Expedited delivery to destinations in North America is \$48; elsewhere \$144. For paper delivery a late charge of 10% of the subscription price will be imposed upon orders received from nonmembers after January 1 of the subscription year.

Back number information. For back issues see www.ams.org/bookstore.

Subscriptions and orders should be addressed to the American Mathematical Society, P.O. Box 845904, Boston, MA 02284-5904 USA. *All orders must be accompanied by payment.* Other correspondence should be addressed to 201 Charles Street, Providence, RI 02904-2294 USA.

Copying and reprinting. Material in this journal may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. Requests can also be made by e-mail to reprint-permission@ams.org.

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

Transactions of the American Mathematical Society is published monthly by the American Mathematical Society at 201 Charles Street, Providence, RI 02904-2294 USA. Periodicals postage is paid at Providence, Rhode Island. Postmaster: Send address changes to *Transactions*, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

© 2003 by the American Mathematical Society. All rights reserved.

This journal is indexed in *Mathematical Reviews*, *Zentralblatt MATH*, *Science Citation Index*®, *Science Citation Index*™-Expanded, *ISI Alerting Services*™, *CompuMath Citation Index*®, and *Current Contents*®/Physical, Chemical & Earth Sciences.

Printed in the United States of America.

⊗ The paper used in this journal is acid-free and falls within the guidelines established to ensure permanence and durability.

TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY

CONTENTS

Vol. 355, No. 8

Whole No. 819

August 2003

Robert Lauter and Sergiu Moroianu , Homology of pseudodifferential operators on manifolds with fibered cusps	3009
Yijun Hu and Tzong-Yow Lee , Moderate deviation principles for trajectories of sums of independent Banach space valued random variables	3047
Yanick Heurteaux , Weierstrass functions with random phases	3065
Stéphane Louboutin , Explicit lower bounds for residues at $s = 1$ of Dedekind zeta functions and relative class numbers of CM-fields	3079
Sophie Huczynska and Stephen D. Cohen , Primitive free cubics with specified norm and trace	3099
David Wright and Wenhua Zhao , D-log and formal flow for analytic isomorphisms of n -space	3117
Nobuo Hara and Ken-ichi Yoshida , A generalization of tight closure and multiplier ideals	3143
Jian Kong , Seshadri constants on Jacobian of curves	3175
Rajesh S. Kulkarni , On the Clifford algebra of a binary form	3181
Jaya N. Iyer , Projective normality of abelian varieties	3209
V. Braungardt and D. Kotschick , Clustering of critical points in Lefschetz fibrations and the symplectic Szpiro inequality	3217
Christian Wolf , On measures of maximal and full dimension for polynomial automorphisms of \mathbb{C}^2	3227
Mark Pollicott , Hausdorff dimension and asymptotic cycles	3241
Marius Dadarlat and Erik Guentner , Constructions preserving Hilbert space uniform embeddability of discrete groups	3253
Jaroslav Tišer , Vitali covering theorem in Hilbert space	3277
George B. Seligman , On idempotents in reduced enveloping algebras ...	3291
Hartmut Logemann, Richard Rebarber, and Stuart Townley , Stability of infinite-dimensional sampled-data systems	3301
Deguang Han , Approximations for Gabor and wavelet frames	3329
Tommaso Pacini , Mean curvature flow, orbits, moment maps	3343
Alexandru D. Ionescu , Singular integrals on symmetric spaces, II	3359
Sergey Antonyan , West's problem on equivariant hyperspaces and Banach-Mazur compacta	3379
Giuseppe de Donno and Alessandro Oliaro , Local solvability and hypoellipticity for semilinear anisotropic partial differential equations	3405



HOMOLOGY OF PSEUDODIFFERENTIAL OPERATORS ON MANIFOLDS WITH FIBERED CUSPS

ROBERT LAUTER AND SERGIU MOROIANU

ABSTRACT. The Hochschild homology of the algebra of pseudodifferential operators on a manifold with fibered cusps, introduced by Mazzeo and Melrose, is studied and computed using the approach of Brylinski and Getzler. One of the main technical tools is a new convergence criterion for tri-filtered half-plane spectral sequences. Using trace-like functionals that generate the 0-dimensional Hochschild cohomology groups, the index of a fully elliptic fibered cusp operator is expressed as the sum of a local contribution of Atiyah-Singer type and a global term on the boundary. We announce a result relating this boundary term to the adiabatic limit of the eta invariant in a particular case.

1. INTRODUCTION

Let X be a compact manifold whose boundary ∂X is the total space of a locally trivial fibration $\varphi : \partial X \rightarrow Y$ of closed manifolds. In contrast to the case of closed manifolds, a great number of algebras of pseudodifferential operators can naturally be associated to this geometric situation. To list a few of them, recall that Richard Melrose introduced a concept of geometric micro-localization that associates to certain classes of Lie algebras \mathcal{V} of vector fields on X algebras of pseudodifferential operators; the Lie algebra \mathcal{V} is then also called a *boundary fibration structure* [26], [30]. The corresponding pseudodifferential calculus $\Psi_{\mathcal{V}}^*(X)$ contains the universal enveloping algebra $\text{Diff}_{\mathcal{V}}^*(X)$ of \mathcal{V} , the \mathcal{V} -*differential operators*, as a subalgebra, and \mathcal{V} -*elliptic* (pseudo)differential operators (a notion that has a natural meaning in the context of boundary fibration structures) can be inverted within $\Psi_{\mathcal{V}}^*(X)$ up to operators of order $-\infty$. Possible candidates for boundary fibration structures on our manifold X are, for instance,

$$\begin{aligned}
 (1) \quad \mathcal{V}_e(X) &:= \{V \in \mathcal{C}^\infty(X, TX) : V \text{ is} \\
 &\quad \text{tangent to the fibers of } \pi \text{ at } \partial X\} \quad (\text{edge-structure}), \\
 \mathcal{V}_{de}(X) &:= \varrho_N \mathcal{V}_e(X) \quad (\text{double-edge structure}), \\
 \mathcal{V}_{\ell e}(X) &:= \varrho_N^\ell \mathcal{V}_e(X) \quad (\ell\text{-fold-edge structure}), \\
 \mathcal{V}_\Phi(X) &:= \{V \in \mathcal{V}_e(X) : V \varrho_N \in \varrho_N^2 \mathcal{C}^\infty(X)\} \quad (\text{fibered-cusp structure}),
 \end{aligned}$$

where $\varrho_N : X \rightarrow \overline{\mathbb{R}}_+$ is a defining function for ∂X , i.e., $\partial X = \{\varrho_N = 0\}$ and $d\varrho_N$ does not vanish on ∂X . Note that the fibered cusp structure $\mathcal{V}_\Phi(X)$ depends slightly on the choice of the boundary defining function. This dependence is discussed in

Received by the editors July 15, 2002 and, in revised form, January 16, 2003.

2000 *Mathematics Subject Classification.* Primary 58J42, 58J20.

Moroianu was partially supported by a DFG-grant (436-RUM 17/7/01) and by the European Commission RTN HPRN-CT-1999-00118 *Geometric Analysis*.

more detail in [23]. Corresponding pseudodifferential calculi were constructed and studied in [22] (edge-structure), [17], [32] (double-edge structure), and [23], [31] (fibered-cusp structure). A pseudodifferential calculus for the ℓ -fold edge structure has been considered so far only for the very special case $\pi = \text{id} : \partial X \rightarrow Y$ and $\ell = 2$, under the name *quadratic scattering structure* [46]. The reader can easily invent more Lie algebras that might introduce boundary fibration structures, but one should be warned that it is not at all straightforward (and sometimes even impossible) to establish the additional properties that ensure the existence of a pseudodifferential calculus. For more details, we refer to [26], [32] and the forthcoming book [25].

It is worth noting that by integrating appropriate Lie algebroids as Nistor did in [39] and using general groupoid techniques, (slightly smaller) pseudodifferential calculi associated to many interesting boundary fibration structures can be constructed simultaneously. We refer to the survey [18] for many examples of this construction.

To give an idea of how the different boundary fibration structures in (1) look locally, we use coordinates $(x, y, z) \in \overline{\mathbb{R}}_+ \times \mathbb{R}_y^n \times \mathbb{R}_z^m$ in a local product decomposition near the boundary, where x is the restriction of the boundary defining function ϱ_N , y is a set of variables on the base Y lifted through φ , and z are variables in the fiber. Then any vector field in one of the boundary fibration structures from (1) can be written locally as a linear combination over $C^\infty(X)$ of the following basic vector fields:

$$(2) \quad \begin{array}{llll} x\partial_x, & x\partial_y, & \partial_z & \text{(edge-structure),} \\ x^2\partial_x, & x^2\partial_y, & x\partial_z & \text{(double-edge structure),} \\ x^{\ell+1}\partial_x, & x^{\ell+1}\partial_y, & x^\ell\partial_z & \text{(\ell-fold-edge structure),} \\ x^2\partial_x, & x\partial_y, & \partial_z & \text{(fibered-cusp structure).} \end{array}$$

In the present paper we compute the Hochschild homology and study its relation to index problems for the fibered cusp calculus, i.e., the pseudodifferential calculus associated to the fibered cusp structure. Starting with the work of Wodzicki [43], [44], [45] and Brylinski and Getzler [3], [4] on Hochschild and cyclic homology of (pseudo)differential operators and on non-commutative residues and Euler classes, homology of pseudodifferential operators has attracted much attention, not only because of its relation to index problems as explained for instance in [34] or [15], [16], but also because of its connections to the non-commutative geometry of Alain Connes [6]. Indeed, algebras of pseudodifferential operators are non-commutative algebras naturally associated to (singular) geometric situations (boundary fibration structures), and there are good reasons to expect that the study of invariants of these non-commutative algebras will reveal information about the geometry and help us to understand what a non-commutative manifold (possibly with singularities) is supposed to be [7].

The interest in the fibered cusp calculus originally introduced in [23], [31] has grown considerably since Nye and Singer [38] used this pseudodifferential calculus to prove an L^2 -index theorem for Dirac operators on $S^1 \times \mathbb{R}^3$, and Vaillant [42] applied the fibered cusp calculus to study spectral and index theory for Dirac operators on manifolds with generalized fibered cusps that occur for instance when compactifying rank one locally symmetric spaces [36].

From (2), we see that the fibered cusp structure interpolates between two interesting boundary fibration structures on manifolds with boundary, namely the

scattering structure corresponding to the trivial fibration $\varphi = \text{id} : \partial X \rightarrow Y = \partial X$, locally given by $x^2\partial_x$ and $x\partial_y$, and the *cuspidal structure* corresponding to $\varphi : \partial X \rightarrow Y = \{\text{pt}\}$ locally generated by $x^2\partial_x$ and ∂_y . For our purposes it is interesting to note that the scattering structure is also a special case of the double-edge structure with $\varphi : \partial X \rightarrow Y = \{\text{pt}\}$. The Hochschild homology for various algebras associated to the cusp algebra has been computed in [34], whereas the Hochschild homology for the analogous algebras in the scattering setting can be extracted as special cases from the homology of the double-edge calculus in [14]. We do not know any way to combine these results to get the Hochschild homology for the fibered cusp calculus directly; nevertheless, the results mentioned above can be used to double-check the computations of this paper. On the other hand, when comparing the results for the double-edge calculus with those for the fibered cusp calculus, we see that Hochschild homology groups yield a functional-analytic way of distinguishing between the two calculi.

The fibered cusp calculus

$$\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X) = \bigcup_{k,m \in \mathbb{Z}} \Psi_{\Phi}^{m,k}(X)$$

is naturally a bifiltered algebra, with the first filtration (m) given by the symbolic order of the pseudodifferential operator whereas the second filtration (k) corresponds to the (negative of the) order of vanishing at the boundary, more precisely at the Φ -front face, a notion that is explained in Section 2.

The algebra $\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X)$ has three interesting ideals, namely

$$\begin{aligned} \Psi_{\Phi}^{-\infty,\mathbb{Z}}(X) &:= \bigcup_{k \in \mathbb{Z}} \bigcap_{m \in \mathbb{Z}} \Psi_{\Phi}^{m,k}(X), \\ \Psi_{\Phi}^{\mathbb{Z},-\infty}(X) &:= \bigcup_{m \in \mathbb{Z}} \bigcap_{k \in \mathbb{Z}} \Psi_{\Phi}^{m,k}(X), \\ \Psi_{\Phi}^{-\infty,-\infty}(X) &:= \bigcap_{k \in \mathbb{Z}} \bigcap_{m \in \mathbb{Z}} \Psi_{\Phi}^{m,k}(X). \end{aligned}$$

The residual ideal $\Psi_{\Phi}^{-\infty,-\infty}$ corresponds to operators with Schwartz kernels in the space of all smooth half-densities that vanish to infinite order at the boundary; it is easily seen to have Hochschild homology only in dimension 0, where it is one-dimensional, the isomorphism to \mathbb{C} being given by the usual operator trace [34]. We are not going to stress this ideal any further. Following [34], let us denote the following quotients by

$$\begin{aligned} \Phi\mathcal{I}_{\partial} &:= \Psi_{\Phi}^{-\infty,\mathbb{Z}} / \Psi_{\Phi}^{-\infty,-\infty}, \\ \Phi\mathcal{A}_{\partial} &:= \Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}} / \Psi_{\Phi}^{\mathbb{Z},-\infty}, \\ \Phi\mathcal{I}_{\sigma} &:= \Psi_{\Phi}^{\mathbb{Z},-\infty} / \Psi_{\Phi}^{-\infty,-\infty}, \\ \Phi\mathcal{A}_{\sigma} &:= \Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}} / \Psi_{\Phi}^{-\infty,\mathbb{Z}}, \\ \Phi\mathcal{A}_{\partial,\sigma} &:= \Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X) / \left(\Psi_{\Phi}^{-\infty,\mathbb{Z}} + \Psi_{\Phi}^{\mathbb{Z},-\infty} \right). \end{aligned}$$

The following diagram summarizes the situation, where the horizontal and vertical sequences are exact:

(3)

$$\begin{array}{ccccccc} & & 0 & & 0 & & 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & \Psi_{\Phi}^{-\infty,-\infty}(X) & \longrightarrow & \Psi_{\Phi}^{\mathbb{Z},-\infty}(X) & \longrightarrow & \Phi\mathcal{I}_{\sigma} \longrightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & \Psi_{\Phi}^{-\infty,\mathbb{Z}}(X) & \longrightarrow & \Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X) & \longrightarrow & \Phi\mathcal{A}_{\sigma} \longrightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & \Phi\mathcal{I}_{\partial} & \longrightarrow & \Phi\mathcal{A}_{\partial} & \longrightarrow & \Phi\mathcal{A}_{\partial,\sigma} \longrightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ & & 0 & & 0 & & 0 \end{array}$$

We are going to compute the Hochschild homology of the ideals $\Phi\mathcal{I}_{\sigma}$ and $\Phi\mathcal{I}_{\partial}$ as well as of the quotients $\Phi\mathcal{A}_{\sigma}$, $\Phi\mathcal{A}_{\partial}$, and $\Phi\mathcal{A}_{\partial,\sigma}$. Moreover, we give geometric descriptions of the long exact sequences in Hochschild homology arising from the six short exact sequences (three rows and three columns) in (3). Not surprisingly, for the computation of the homology of the symbolic algebras $\Phi\mathcal{I}_{\sigma}$, $\Phi\mathcal{A}_{\sigma}$, and $\Phi\mathcal{A}_{\partial,\sigma}$ we use the same approach as Brylinski and Getzler in [4], and obtain, for $k \in \mathbb{N}_0$ and with $N := \dim X$,

$$\begin{aligned} HH_k(\Phi\mathcal{I}_{\sigma}) &\cong H_{rel}^{2N-k}(\Phi S^*X \times S^1), \\ HH_k(\Phi\mathcal{A}_{\sigma}) &\cong H_{abs}^{2N-k}(\Phi S^*X \times S^1) \oplus H^{2N-1-k}(\Phi S^*X|_{\partial X} \times S^1), \\ HH_k(\Phi\mathcal{A}_{\partial,\sigma}) &\cong H^{2N-k}(\Phi S^*X|_{\partial X} \times S^1 \times S^1), \end{aligned}$$

where H_{rel}^* (resp. H_{abs}^*) stand for the relative (resp. absolute) de Rham cohomology of a compact manifold with boundary. The fibered cusp cosphere bundle ΦS^*X shows up naturally in the analysis of the fibered cusp calculus and is in fact diffeomorphic to the usual cosphere bundle S^*X , though not naturally. However, even though the idea of the computation is the same as in [4], namely to use an appropriate form of symplectic duality to identify the d^1 in the spectral sequence, convergence of the spectral sequence is an issue that is often neglected, and we spend some time to give the complete argument. In fact, we are naturally led to consider tri-filtered differential complexes, where the filtration indices run in \mathbb{Z} . In this setting convergence issues become very complicated, although it is rather satisfactory that they can be affirmatively answered. We give a general criterion for convergence which might well be applied in other situations.

More surprising is the computation of the homology of the boundary ideal $\Phi\mathcal{I}_{\partial}$ and the boundary algebra $\Phi\mathcal{A}_{\partial}$. We use a non-canonical morphism of algebras $\Theta : \Psi_{sc,cpt}^{-\infty}(Y_{\epsilon}) \rightarrow \Psi_{\Phi}^{-\infty}(X)$ of the algebra of (compactly supported) scattering pseudodifferential operators of order $-\infty$ on the cylinder $Y_{\epsilon} := Y \times [0, \epsilon)$ to the smoothing fibered cusp operators on X to reduce the computation of the Hochschild homology of $\Phi\mathcal{I}_{\partial}$ to that of the smoothing boundary ideal ${}^s\mathcal{I}_{\partial}$ of the scattering

calculus on Y_ε . More precisely, we have

$$HH_k(\Phi\mathcal{I}_\partial) \cong H^{n+1-k}(Y) \oplus H^{n-k}(Y),$$

where, for simplicity, we have assumed that the base Y of the fibration is orientable; the general case is obtained in Theorem 4.6 using cohomology with coefficients in the orientation bundle. The homology of the fibered cusp boundary algebra follows then from a long exact sequence that we describe explicitly in geometric terms (Proposition 6.1, Theorem 6.2).

Starting from a different approach using differentiable groupoids, Benamèur and Nistor [2] developed a very general machinery that computes, when specializing to our setting, the E^∞ term in the spectral sequence of the Hochschild homology of the symbolic algebra $\Phi\mathcal{A}_\sigma$. Finally, note that computing Hochschild homology for algebras of pseudodifferential operators is not restricted to the algebras mentioned above: for instance, the Hochschild homology for Boutet de Monvel's algebra on a compact manifold with boundary has been computed by Nest and Schrohe [37]; the second author has computed the homology for the adiabatic limit algebra [35].

The second part of the paper is devoted to residue functionals, traces and index formulas for the fibered cusp calculus. As in [14] or [34], we introduce several linear functionals on $\Psi_\Phi^{\mathbb{Z},\mathbb{Z}}(X)$ that descend to traces on some of the ideals and quotients considered above by taking appropriate coefficients in the Laurent expansion of the meromorphic extension $Z_{\partial N,Q}(A)$ of the double-zeta function $(\lambda, z) \mapsto \text{Tr}(A\varrho_N^z Q^{-\lambda})$; here, $Q \in \Psi_\Phi^{1,0}(X)$ is a fixed elliptic, symmetric, strictly positive operator. Using these functionals, we can give a formula for the index of Fredholm operators A in the fibered cusp calculus that extends the Atiyah-Patodi-Singer formula for closed manifolds in the form [34] to the fibered cusp calculus:

$$\text{index}(A) = \overline{AS}(A) - \lim_a \eta(A)/2.$$

Here $\overline{AS}(A)$ is given in terms of Hochschild homological functionals that depend only on the symbolic behavior of the operator A but still involve an inverse of A up to trace class remainders, while $\lim_a \eta(A)$ is a term that depends only on the behavior of A near the boundary. The term $\overline{AS}(A)$ can be linked to the asymptotics of the heat kernel [13]. The notation $\lim_a \eta(A)$ is motivated by the fact that in the special case $Y = S^1$ and A a differential operator with a Dirac-type decomposition near the boundary, the non-local boundary contribution to the index can be identified as the adiabatic limit of the eta invariant of the “restriction to the boundary” of A . This result is developed in [13].

The paper is organized as follows: In Section 2 we review some basic facts about the fibered cusp calculus; Section 3 is devoted to the construction of the morphism $\Theta : \Psi_{sc,cpt}^{-\infty}(Y_\varepsilon) \rightarrow \Psi_\Phi^{-\infty}(X)$, which is used in Section 4 to define a generalization of the Hochschild-Kostant-Rosenberg map in order to determine the Hochschild homology of $\Phi\mathcal{I}_\partial$. The computation of the Hochschild homology of the symbol algebras $\Phi\mathcal{I}_\sigma$, $\Phi\mathcal{A}_\sigma$, and $\Phi\mathcal{A}_{\partial,\sigma}$ can be found in Section 5. The long exact sequence in Hochschild homology arising from the boundary sequence

$$0 \longrightarrow \Phi\mathcal{I}_\partial \longrightarrow \Phi\mathcal{A}_\partial \longrightarrow \Phi\mathcal{A}_{\partial,\sigma} \longrightarrow 0$$

is described in Section 6 in geometric terms. In Section 7 we introduce and study various trace-like functionals on the fibered cusp calculus that generate the 0-dimensional Hochschild homology groups. As in [16], [14], [34], these functionals

can be used to derive an index formula for fully elliptic fibered cusp operators. The formula is presented in Section 8.

The paper contains two appendices. In Appendix A, for the convenience of the reader, we review the definition of Schwartz (resp. symbolic) sections of vector bundles. Appendix B contains the definition of the Hochschild chain spaces for topologically filtered algebras in the sense of [2], and a general criterion for the convergence of the spectral sequence associated to a tri-filtered complex, which is applied in the body of the paper to the Hochschild complexes of the symbol and boundary Φ -algebras.

As usual, we write $S^{[m]}(V)$ for the space of all smooth functions $V \setminus \{0\} \rightarrow \mathbb{C}$ of a vector bundle V that are positively homogeneous of degree $m \in \mathbb{C}$.

ACKNOWLEDGMENTS

We are indebted to Richard Melrose and Victor Nistor for explaining to us their results in [34], which greatly influenced this paper, and to an anonymous referee for a careful reading of the manuscript. The first named author would like to thank Michael Singer and Boris Vaillant for some private lessons on the Φ -calculus.

2. REVIEW OF THE FIBERED CUSP CALCULUS

In this section we review basic properties of the fibered cusp calculus. For details and most of the proofs we refer to [23], [31], and [42].

The fibered cusp structure space. Throughout this paper, $X = X^N$ stands for a smooth, compact manifold with boundary ∂X that is the total space of a locally trivial fibration $\varphi : \partial X \rightarrow Y = Y^n$ of closed manifolds. For simplicity, let us assume that ∂X is connected; so the fibers of φ are all diffeomorphic, say to a closed manifold $F = F^m$. For the dimension N of X we thus have $N = n + m + 1$. Furthermore, we fix a boundary defining function $\varrho_N : X \rightarrow \overline{\mathbb{R}}_+$ for ∂X , i.e., ϱ_N is a smooth function such that $\partial X = \{\varrho_N = 0\}$ and $d\varrho_N \neq 0$ at ∂X .

Let $\mathcal{V}_e(X)$ be the Lie algebra of all *edge vector fields*, i.e., all smooth vector fields on X that are tangent to the fibers of φ at the boundary [22]. Then the vector fields in

$$\mathcal{V}_\Phi(X) := \{V \in \mathcal{V}_e(X) : V\varrho_N \in \varrho_N^2 C^\infty(X)\}$$

are called *fibered cusp* or simply Φ -*vector fields*. Note that the definition of the Lie algebra $\mathcal{V}_\Phi(X)$ depends slightly on the choice of the boundary defining function ϱ_N [23]. To give a local description of Φ -vector fields near the boundary, let

(4)
$$(x, y, z) : X \supseteq U \longrightarrow \overline{\mathbb{R}}_+ \times \mathbb{R}_y^n \times \mathbb{R}_z^m$$

be coordinates of a local product decomposition near the boundary, i.e., $x = \varrho_N|_U$, y are variables on the base Y lifted through φ , and with z corresponding to coordinates on the fiber F . Then $V \in \mathcal{V}_\Phi(X)$ is of the form

$$V(x, y, z) = a(x, y, z)x^2\partial_x + \sum_{j=1}^n b_j(x, y, z)x\partial_{y_j} + \sum_{k=1}^m c_k(x, y, z)\partial_{z_k}, \quad (x, y, z) \in U$$

with coefficients a, b_j, c_k smooth up to $x = 0$. Thus, by the Serre-Swan theorem there exists a smooth vector bundle ${}^\Phi TX \rightarrow X$ together with a map $\iota_\Phi : {}^\Phi TX \rightarrow TX$ of vector bundles over X such that

$$\iota_\Phi(C^\infty(X, {}^\Phi TX)) = \mathcal{V}_\Phi(X).$$

We call ${}^{\Phi}TX$ the Φ -tangent bundle, and its dual ${}^{\Phi}T^*X$ the Φ -cotangent bundle. Note that the compactness of X is not necessary for the existence of the Φ -tangent bundle ${}^{\Phi}TX$. In the special case when the fibration $\varphi = \text{id} : \partial X = Y \rightarrow Y$ is trivial, Φ -vector fields are also known as *scattering vector fields* [28]; accordingly, in that case the Φ -tangent bundle is called the *scattering tangent bundle*, and we write ${}^{\Phi}TX = {}^{sc}TX$. By the very definition, the restriction of the canonical map $\iota_{\Phi} : {}^{\Phi}TX \rightarrow TX$ and its dual $\iota_{\Phi}^* : T^*X \rightarrow {}^{\Phi}T^*X$ to the interior $X_0 := X \setminus \partial X$ are isomorphisms. Thus, the canonical symplectic form ω on $T^*X|_{X_0}$ can be pushed forward to a singular closed 2-form ω_{Φ} on ${}^{\Phi}T^*X|_{X_0}$ giving ${}^{\Phi}T^*X|_{X_0}$ the structure of a symplectic manifold. With respect to local coordinates (x, y, z) as in (4), and the associated local coordinates $(x, y, z, \xi, \eta, \zeta)$ on ${}^{\Phi}T^*X|_U$, ω_{Φ} is given by

$$\omega_{\Phi} = \frac{dx}{x^2} \wedge d\xi + \frac{dy}{x} \wedge d\eta + dz \wedge d\zeta + x\eta \frac{dx}{x^2} \wedge \frac{dy}{x};$$

so $\omega_{\Phi} \in \mathcal{C}^{\infty}({}^{\Phi}T^*X, \Lambda^2({}^{\Phi}T^*({}^{\Phi}T^*X)))$. The associated Φ -Poisson bracket $\{f, g\}_{\Phi}$ is locally given by

$$(5) \quad \begin{aligned} \{f, g\}_{\Phi} = & \quad x^2 \partial_x f \partial_{\xi} g - x^2 \partial_x g \partial_{\xi} f + x \partial_y f \partial_{\eta} g - x \partial_y g \partial_{\eta} f \\ & + \partial_z f \partial_{\zeta} g - \partial_z g \partial_{\zeta} f + x (\eta \partial_{\eta} f \partial_{\xi} g - \eta \partial_{\eta} g \partial_{\xi} f). \end{aligned}$$

Let us denote the kernel of the map $\iota_{\Phi}|_{\partial X} : {}^{\Phi}TX|_{\partial X} \rightarrow TX|_{\partial X}$ by ${}^{\Phi}N\partial X$. Note that ${}^{\Phi}N\partial X$ is locally generated by the vector fields $x^2 \partial_x$ and $x \partial_y$; hence, ${}^{\Phi}N\partial X$ is a vector bundle over ∂X . Let $V\partial X \rightarrow \partial X$ be the vertical tangent bundle, i.e., the kernel of the differential $T\varphi : T\partial X \rightarrow TY$. Then we have the following short exact sequence of vector bundles over ∂X :

$$(6) \quad 0 \longrightarrow {}^{\Phi}N\partial X \longrightarrow {}^{\Phi}TX \longrightarrow V\partial X \longrightarrow 0.$$

In fact, the bundle ${}^{\Phi}N\partial X$ is the pull-back $\varphi^*({}^{\Phi}NY)$ of a vector bundle ${}^{\Phi}NY$ over Y . Since this bundle plays an important role in our computations, let us recall one way of constructing it. For small $\varepsilon > 0$, let $Y_{\varepsilon} := Y \times [0, \varepsilon)$, and let ${}^{sc}TY_{\varepsilon}$ be the corresponding scattering tangent bundle. Then ${}^{\Phi}NY := {}^{sc}TY_{\varepsilon}|_{Y \times \{0\}}$ has the desired properties. Indeed, with respect to the local coordinates (4), the map $\vartheta : {}^{\Phi}N\partial X \rightarrow {}^{\Phi}NY$ with $\vartheta(x^2 \partial_x|_{(0, y, z)}) = x^2 \partial_x|_{(y, 0)}$ and $\vartheta(x \partial_y|_{(0, y, z)}) = x \partial_y|_{(y, 0)}$ is well-defined, does not depend on the choice of local coordinates, and makes the following pull-back diagram commutative.

$$\begin{array}{ccc} {}^{\Phi}N\partial X & \xrightarrow{\vartheta} & {}^{\Phi}NY \\ \downarrow & & \downarrow \\ \partial X & \xrightarrow{\varphi} & Y \end{array}$$

The fibered cusp double space. For the definition of the fibered cusp calculus we need a compact manifold with corners that is obtained by blowing up a sequence of p -submanifolds of the product X^2 . For the concept of blowing up p -submanifolds we refer for instance to [10], [28] and the forthcoming book [25].

Let $\beta_b^2 : X_b^2 := [X^2; (\partial X)^2] \rightarrow X^2$ be the b -blow-up [27]. By [27, Lemma 4.1] the b -front face ff_b , i.e., the new boundary face of X_b^2 obtained by the b -blow-up, is canonically diffeomorphic to the product $[-1, 1] \times \partial X \times \partial X$. Note that we have fixed a boundary defining function ϱ_N to define the fibered cusp structure. Then

$$B_{\Phi} := \{(0, q, q') \in \text{ff}_b = [-1, 1] \times \partial X \times \partial X : \varphi(q) = \varphi(q')\}$$

is a p -submanifold of X_b^2 , and

$$\beta_\Phi^2 : X_\Phi^2 := [X_b^2; B_\Phi] \xrightarrow{\beta} X_b^2 \xrightarrow{\beta_b^2} X^2$$

is called the *fibered cusp* or briefly the Φ -double space. It is a compact manifold with corners up to codimension two. Let $\Delta \subseteq X^2$ be the diagonal. Then the submanifold $\Delta_\Phi := (\beta_\Phi^2)^{-1}(\Delta \setminus (\partial X)^2)^{X_\Phi^2}$ is said to be the *lifted diagonal*, and the boundary face ff_Φ of X_Φ^2 that meets the diagonal is called the *fibered cusp* or simply Φ -front face. The Φ -front face is of great importance for understanding the properties of the fibered cusp calculus near the boundary. As explained in [23], [42], ff_Φ is canonically diffeomorphic to

$$(7) \quad \text{ff}_\Phi \cong \partial X \times_Y {}^\Phi \overline{N} \partial X = \partial X \times_Y \partial X \times_Y {}^\Phi \overline{N} Y,$$

where ${}^\Phi \overline{N} \partial X$ (resp. ${}^\Phi \overline{N} Y$) is the radial compactification of ${}^\Phi N \partial X$ (resp. ${}^\Phi N Y$) as explained in Appendix A. Consequently, the interior $\text{ff}_\Phi^{\text{int}}$ of ff_Φ is canonically diffeomorphic to $\partial X \times_Y {}^\Phi N \partial X = \partial X \times_Y \partial X \times_Y {}^\Phi N Y$. If necessary, we use local coordinates

$$(8) \quad r = x + x', \quad S = \frac{x - x'}{(x + x')^2}, \quad U = \frac{y - y'}{x + x'}, \quad y', z, z'$$

near $\text{ff}_\Phi^{\text{int}} = \{r = 0\}$. The lifted diagonal is then given by $\Delta_\Phi = \{S = 0, U = 0, z = z'\}$. With respect to the decomposition (7) of the Φ -front face ff_Φ , the variables $S \in \mathbb{R}_S$ and $U \in \mathbb{R}_U^n$ then correspond to the linear variables in the fibers of ${}^\Phi N Y$.

The Φ -density bundle. Finally, for a density bundle that is adapted to the fibered cusp calculus, apply the smooth functor Ω^α of α -densities to the Φ -cotangent bundle ${}^\Phi T^* X$ to obtain the bundle ${}^\Phi \Omega^\alpha(X)$ of Φ - α -densities. The choice of local product coordinates as in (4) trivializes ${}^\Phi \Omega^\alpha(X)|_U$. A non-vanishing section is given by $|\frac{dx}{x^{n+2}} dy dz|^\alpha$.

On the Φ -double space X_Φ^2 , we use the Φ -kernel-half density bundle $KD_\Phi^{1/2} := \varrho_{\text{ff}_\Phi}^{-(2+n)/2} \Omega^{1/2}(X_\Phi^2)$. It is completely characterized by the space of its \mathcal{C}^∞ -sections [27, Lemma 8.6], namely

$$\mathcal{C}^\infty(X_\Phi^2, KD_\Phi^{1/2}) = \varrho_{\text{ff}_\Phi}^{-(2+n)/2} \mathcal{C}^\infty(X_\Phi^2, \Omega^{1/2}(X_\Phi^2)).$$

Here, as usual, $\varrho_{\text{ff}_\Phi} : X_\Phi^2 \rightarrow \overline{\mathbb{R}}_+$ stands for a defining function of the Φ -front face ff_Φ .

The fibered cusp calculus. The fibered cusp calculus is defined by characterizing the singularities of the Schwartz kernel associated to a bounded linear operator $A : \dot{\mathcal{C}}^\infty(X, {}^\Phi \Omega^{1/2}) \rightarrow \mathcal{C}^{-\infty}(X, {}^\Phi \Omega^{1/2})$. A convenient description of the Schwartz kernel $k_A \in \mathcal{C}^{-\infty}(X^2, {}^\Phi \Omega^{1/2} \boxtimes {}^\Phi \Omega^{1/2})$ is possible when replacing the double X^2 with the Φ -double space X_Φ^2 . Recall that the blow-down map $\beta_\Phi^2 : X_\Phi^2 \rightarrow X^2$ induces via pull-back and duality an isomorphism

$$(\beta_\Phi^2)_* : \mathcal{C}^{-\infty}(X_\Phi^2, KD_\Phi^{1/2}) \rightarrow \mathcal{C}^{-\infty}(X^2, {}^\Phi \Omega^{1/2} \boxtimes {}^\Phi \Omega^{1/2}).$$

We call the image κ_A of k_A under this isomorphism the *lifted Schwartz kernel* of A .

Definition 2.1. The space $\Psi_\Phi^m(X)$ of *fibered cusp* or briefly Φ -operators of order $m \in \mathbb{C}$ consists of all continuous linear operators

$$A : \dot{\mathcal{C}}^\infty(X, {}^\Phi \Omega^{1/2}) \rightarrow \mathcal{C}^{-\infty}(X, {}^\Phi \Omega^{1/2})$$

whose lifted Schwartz kernel κ_A belongs to the space $I_{cl}^m(X_\Phi^2, \Delta_\Phi; KD_\Phi^{1/2})$ of classically conormal distributions that vanish to infinite order at all boundary faces of X_Φ^2 other than the front face ff_Φ , and are extendible across ff_Φ .

A fibered cusp operator $A \in \Psi_\Phi^m(X)$ maps $\dot{C}^\infty(X, \Phi\Omega^{1/2})$ continuously into $\dot{C}^\infty(X, \Phi\Omega^{1/2})$; thus, composition of operators as well as conjugation with arbitrary complex powers of the boundary defining function ϱ_N are well-defined. Since we get the lifted Schwartz kernel of $\varrho_N^z A \varrho_N^{-z}$ by multiplying the lifted Schwartz kernel of $A \in \Psi_\Phi^m(X)$ by a smooth function on the interior of X_Φ^2 that extends smoothly to the front face, is identical to 1 at Δ_Φ and polynomially bounded near all boundary faces other than the front face, we have $\varrho_N^z \Psi_\Phi^m(X) \varrho_N^{-z} = \Psi_\Phi^m(X)$, and we can define for $k \in \mathbb{C}$

$$\Psi_\Phi^{m,k}(X) := \varrho_N^{-k} \Psi_\Phi^m(X) = \Psi_\Phi^m(X) \varrho_N^{-k}.$$

Composition of operators leads to a bilinear map

$$\Psi_\Phi^{m,k}(X) \times \Psi_\Phi^{m',k'}(X) \rightarrow \Psi_\Phi^{m+m',k+k'}(X)$$

that defines on $\Psi_\Phi^{\mathbb{Z},\mathbb{Z}}(X) := \bigcup_{m \in \mathbb{Z}} \bigcup_{k \in \mathbb{Z}} \Psi_\Phi^{m,k}(X)$ the bi-filtered algebra structure already considered in the introduction; in the sequel we use the notation introduced therein. We emphasize the negative sign in the definition of the second filtration, whose purpose is to make both filtrations increasing.

The filtration by the order of vanishing at the boundary on the algebra $\Psi_\Phi^{\mathbb{Z},\mathbb{Z}}(X)$ induces in particular filtrations on $\Phi\mathcal{I}_\partial$ and $\Phi\mathcal{A}_\partial$, namely

$$\begin{aligned} \Phi\mathcal{I}_\partial^k &:= \Psi_\Phi^{-\infty,k}(X) / \Psi_\Phi^{-\infty,-\infty}(X), \\ \Phi\mathcal{A}_\partial^{m,k} &:= \Psi_\Phi^{m,k}(X) / \Psi_\Phi^{m,-\infty}(X) \end{aligned}$$

with quotients

$$\begin{aligned} \Phi\mathcal{I}_\partial^{[m]} &:= \Phi\mathcal{I}_\partial^m / \Phi\mathcal{I}_\partial^{m-1} = \Psi_\Phi^{-\infty,m}(X) / \Psi_\Phi^{-\infty,m-1}(X), \\ \Phi\mathcal{A}_\partial^{m,[k]} &:= \Phi\mathcal{A}_\partial^{m,k} / \Phi\mathcal{A}_\partial^{m,k-1} = \Psi_\Phi^{m,k}(X) / \Psi_\Phi^{m,k-1}(X). \end{aligned}$$

As above, we let $\Phi\mathcal{I}_\partial^{[\mathbb{Z}]} := \bigoplus_{m \in \mathbb{Z}} \Phi\mathcal{I}_\partial^{[m]}$, and so on. Similarly, the filtration of $\Psi_\Phi^{\mathbb{Z},\mathbb{Z}}(X)$ by the operator order induces filtrations of $\Phi\mathcal{I}_\sigma$ and $\Phi\mathcal{A}_\sigma$, and we use the corresponding notation without any further comments. The usual symbol map for conormal distributions together with the fact that the density bundle of ΦT^*X is canonically trivialized by the symplectic form leads to the principal symbol map for the fibered cusp calculus,

$$0 \longrightarrow \Psi_\Phi^{m-1,k}(X) \longrightarrow \Psi_\Phi^{m,k}(X) \xrightarrow{\Phi\sigma^{(m,k)}} S^{[m]}(\Phi T^*X) \longrightarrow 0,$$

which is multiplicative in the obvious sense. Let $\Phi\overline{T}^*X$ be the radial compactification of the Φ -cotangent bundle ΦT^*X in the sense of Appendix A. Then the choice of a defining function

$$\varrho_\sigma : \Phi\overline{T}^*X \rightarrow \overline{\mathbb{R}}_+$$

for the Φ -cosphere bundle

$$\Phi S^*X := (\Phi T^*X \setminus \{0\}) / \mathbb{R}_+ \subseteq \Phi\overline{T}^*X$$

yields an identification of the space $S^{[m]}(\Phi T^*X)$ with the space $\mathcal{C}^\infty(\Phi S^*X)$.

The normal homomorphism. In addition to the symbolic behavior at the lifted diagonal, the behavior of the lifted Schwartz kernel κ_A of an operator $A \in \Psi^m_\Phi(X)$ at the Φ -front face determines for instance the Fredholm properties of a fibered cusp operator. So, let us consider $\mathcal{N}_\Phi(A) := \kappa_A|_{\text{ff}_\Phi}$ directly. As shown in [23], the distributional density $\mathcal{N}_\Phi(A)$ can be understood as a family $\mathcal{N}_\Phi(A)(y)$, $y \in Y$, of pseudodifferential operators of order m on the space $\varphi^{-1}(y) \times {}^\Phi N_y Y$ that are translation invariant with respect to the linear variables in ${}^\Phi N_y Y$, and depend smoothly on $y \in Y$. The space of all these operators is denoted by $\Psi^m_{\text{sus}({}^\Phi NY) - \varphi}(\partial X)$. For a precise definition we refer to [23]. The family $\mathcal{N}_\Phi(A)$ is called the *normal operator* of A , and \mathcal{N}_Φ induces a family of short exact sequences

$$(9) \qquad 0 \longrightarrow \varrho_N \Psi^m_\Phi(X) \longrightarrow \Psi^m_\Phi(X) \xrightarrow{\mathcal{N}_\Phi} \Psi^m_{\text{sus}({}^\Phi NY) - \varphi}(\partial X) \longrightarrow 0,$$

which are multiplicative in the obvious way. The normal operator extends to $A \in \Psi^{m,k}_\Phi(X)$ by defining $\mathcal{N}^{(k)}_\Phi(A) := \mathcal{N}_\Phi(\varrho_N^k A) = \mathcal{N}_\Phi(A \varrho_N^k) \in \Psi^m_{\text{sus}({}^\Phi NY) - \varphi}(\partial X)$.

Fourier transform along the fibers of ${}^\Phi NY$, i.e., with respect to the variables S and U in (8), transforms a family $P = p(y, z, D_z, D_S, D_U) \in \Psi^m_{\text{sus}({}^\Phi NY) - \varphi}(\partial X)$ into a family $\tilde{P} = \tilde{p}(y, z, D_z, \xi, \eta)$ of pseudodifferential operators on $\varphi^{-1}(y)$ that depend smoothly on $y \in Y$ and symbolically of order m on the dual variables $(\xi, \eta) \in {}^\Phi N^*_y Y$ to $(S, U) \in {}^\Phi N_y Y$ [31]. Let us write $\tilde{\Psi}^m_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ for the corresponding space of operators, and $\tilde{\mathcal{N}}^{(k)}_\Phi : \Psi^{m,k}_\Phi(X) \rightarrow \tilde{\Psi}^m_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ for the corresponding normal operator. For the sake of completeness, let us give a description of families in $\tilde{\Psi}^M_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ using local data. By a partition of unity in Y we can assume that the family $\tilde{P} \in \tilde{\Psi}^M_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ is compactly supported in an open set V such that $\varphi^{-1}(V) = V \times F$. Then we have, up to a density factor in the y -variable,

$$(10) \qquad \tilde{P} \in \mathcal{C}^\infty_c(V, \Psi^M(F; \mathbb{R}_\xi \times \mathbb{R}^n_\eta)),$$

where $\Psi^M(F; \mathbb{R}_\xi \times \mathbb{R}^n_\eta)$ denotes the space of parameter-dependent or suspended pseudodifferential operators of order M on F in the sense of [40] – see for instance also [9], [19], [21] or [29] for the one-dimensional case. For $M = -\infty$, (10) is equivalent to $\tilde{P} \in \mathcal{C}^\infty_c(V) \hat{\otimes}_\pi S(\mathbb{R}_\xi \times \mathbb{R}^n_\eta) \hat{\otimes}_\pi \Psi^{-\infty}(F)$. For arbitrary $M \in \mathbb{C}$ a partition of unity in F identifies \tilde{P} (up to operators of order $-\infty$) with a finite sum of families of the form $\tilde{p}(y, z, D_z, \xi, \eta)$ for some compactly supported classical symbols $\tilde{p} \in \mathcal{C}^\infty_c(V, S^M_{cl}(\mathbb{R}^m_z, \mathbb{R}^m_\zeta \times \mathbb{R}_\xi \times \mathbb{R}^n_\eta))$.

In the sequel we use these equivalent pictures of the normal operator simultaneously.

Fiberwise trace for smoothing suspended operators. There exists a canonical trace on $\tilde{\Psi}^{-\infty}_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$, which we review here. An extension of this trace to families in $\tilde{\Psi}^m_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ for arbitrary finite $m \in \mathbb{Z}$ is constructed in Section 7. As explained above, elements of $\tilde{\Psi}^{-\infty}_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ are simply smooth sections of the half-density bundle $\Omega^{1/2}(\partial X \times_Y \partial X \times_Y {}^\Phi \overline{N^*Y})$ that vanish to infinite order at the boundary. Thus, inverse Fourier transform along the fibers of ${}^\Phi N^*Y$ transforms an element $\tilde{h} \in \tilde{\Psi}^{-\infty}_{\text{sus}({}^\Phi N^*Y) - \varphi}(\partial X)$ into a smooth section h of

$$\Omega^{1/2}(\partial X \times_Y \partial X \times_Y {}^\Phi \overline{NY}) = \Omega^{1/2}(\text{ff}_\Phi)$$

that vanishes to infinite order at the boundary, i.e., we have $h \in \Psi_{\text{sus}(\Phi NY) - \varphi}^{-\infty}(\partial X)$. Here we have used the identification $\text{ff}_{\Phi} \cong \partial X \times_Y \partial X \times_Y {}^{\Phi}\overline{NY}$ by (7). Recall that the blow-down map $\beta_{\Phi}^2 : X_{\Phi}^2 \rightarrow X^2$ together with the projection $X^2 \rightarrow X$ onto either the left or the right factor identifies $\Delta_{\Phi} \cap \text{ff}_{\Phi}$ with ∂X . On the other hand, $\Delta_{\Phi} \cap \text{ff}_{\Phi}$ corresponds under the diffeomorphism (7) to

$$D_{\Phi} := \{(q, q, v) \in \partial X \times_Y \partial X \times_Y {}^{\Phi}\overline{NY} : \varphi(q) = y \text{ and } v = 0 \in {}^{\Phi}N_y Y\},$$

and we have canonical isomorphisms

$$\begin{aligned} (11) \quad & \Omega^{1/2}(\partial X \times_Y \partial X \times_Y {}^{\Phi}\overline{NY})|_{D_{\Phi}} \\ & \cong \varphi^* \Omega^{1/2}(Y) \otimes \Omega_{\text{fiber}}^{1/2}(\partial X) \otimes \Omega_{\text{fiber}}^{1/2}(\partial X) \otimes \varphi^* \Omega_{\text{fiber}}^{1/2}({}^{\Phi}NY)|_Y \\ & \cong \varphi^* \Omega^{1/2}(Y) \otimes \Omega_{\text{fiber}}^1(\partial X) \otimes \varphi^* \Omega^{1/2}(Y) \\ & \cong \Omega^1(\partial X), \end{aligned}$$

where (11) follows from the fact that the normal direction to the boundary in ${}^{\Phi}NY = {}^{sc}TY_{\varepsilon}|_{Y \times \{0\}}$ leads to a canonical trivial factor in the fiber half-density bundle $\Omega_{\text{fiber}}^{1/2}({}^{\Phi}NY)|_Y$. Thus, we can integrate $h|_{\partial X}$ over $\partial X = \Delta_{\Phi} \cap \text{ff}_{\Phi}$ and obtain a map

$$\overline{\text{Tr}} : \Psi_{\text{sus}(\Phi N^*Y) - \varphi}^{-\infty}(\partial X) \longrightarrow \mathbb{C} : \tilde{h} \longmapsto \int_{\partial X} h|_{\partial X}$$

satisfying

$$\overline{\text{Tr}}(\tilde{h}_1 \tilde{h}_2) = \overline{\text{Tr}}(\tilde{h}_2 \tilde{h}_1)$$

for $\tilde{h}_1 \in \Psi_{\text{sus}(\Phi N^*Y) - \varphi}^{-\infty}(\partial X)$ and $\tilde{h}_2 \in \Psi_{\text{sus}(\Phi N^*Y) - \varphi}^m(\partial X)$. Indeed, by a partition of unity in Y we can assume that the family $\tilde{h} \in \Psi_{\text{sus}(\Phi N^*Y) - \varphi}^{-\infty}(\partial X)$ of pseudodifferential operators is compactly supported in an open set $V \subseteq Y$ such that $\varphi^{-1}(V) \cong V \times F$. Then we have, up to the density factors, $\tilde{h}|_V \in \mathcal{C}_c^{\infty}(V, \Psi^{-\infty}(F; \mathbb{R}_{\xi} \times \mathbb{R}_{\eta}^n))$ and $\overline{\text{Tr}}(\tilde{h}) = \int_V \int_{\mathbb{R}_{\xi}} \int_{\mathbb{R}_{\eta}^n} \text{Tr}(\tilde{h}(y)(\xi, \eta)) d\eta d\xi dy$, where Tr stands for the usual operator trace on operators on the fiber F . In particular, we see that $\overline{\text{Tr}}$ vanishes on commutators because Tr has this property. Moreover, by gluing the local pieces together we obtain the following form of $\overline{\text{Tr}}$:

$$(12) \quad \overline{\text{Tr}} : \Psi_{\text{sus}(\Phi N^*Y) - \varphi}^{-\infty}(\partial X) \longrightarrow \mathbb{C} : \tilde{h} \longmapsto \int_Y \overline{\text{Tr}}_{(y)}(\tilde{h}|_{\varphi^{-1}(y)}),$$

where $\overline{\text{Tr}}_{(y)} : \Psi_{\text{sus}}^{-\infty}(\varphi^{-1}(y); {}^{\Phi}N_y^*Y) \ni \tilde{h} \mapsto \int_{{}^{\Phi}N_y^*Y} \text{Tr}(h(\xi, \eta))$ is the canonical trace on the space of suspended operators on the fiber $\varphi^{-1}(y)$. For the one-suspended case the canonical trace $\overline{\text{Tr}}$ was first considered in [29].

Moreover, by (12) or directly from the definition of $\overline{\text{Tr}}$ we see that $\overline{\text{Tr}}$ extends to $\tilde{\Psi}_{\text{sus}(\Phi N^*Y) - \varphi}^M(\partial X)$ as long as $M < -N = -\dim X$.

Commuting fibered cusp operators with $\log \varrho_N$. Since $A \in \Psi_{\Phi}^{m,k}(X)$ as well as $\log \varrho_N$ act as bounded operators on $\dot{\mathcal{C}}^{\infty}(X, {}^{\Phi}\Omega^{1/2})$, we can consider their commutator $[A, \log \varrho_N] : \dot{\mathcal{C}}^{\infty}(X, {}^{\Phi}\Omega^{1/2}) \rightarrow \dot{\mathcal{C}}^{\infty}(X, {}^{\Phi}\Omega^{1/2})$.

Lemma 2.2. *The commutator with $\log \varrho_N$ yields a map*

$$[\cdot, \log \varrho_N] : \Psi_{\Phi}^{m,k}(X) \longrightarrow \Psi_{\Phi}^{m-1,k-1}(X).$$

Proof. Multiplication with $\log \varrho_N$ is a local operator; hence we can restrict ourselves to operators A with lifted Schwartz kernel κ_A supported in a coordinate patch of X^2_Φ . Let us first assume that this patch is close to the intersection of the lifted diagonal Δ_Φ with the fibered-cusp front face ff_Φ . We use local coordinates $(r, S, U, y'z, z')$ as in (8). A straightforward computation shows that the lifted kernel κ_B of $B := [A, \log \varrho_N]$ is given by $\kappa_B = -\log \left(\frac{1+rS}{1-rS} \right) \kappa_A$. Since $\log \left(\frac{1+rS}{1-rS} \right)$ vanishes to first order at $\Delta_\Phi = \{S = 0, U = 0, z = z'\}$ and $\text{ff}_\Phi = \{r = 0\}$ and is polynomially bounded for $|S| \rightarrow \infty$, we obtain $B \in \Psi^{m-1, k-1}_\Phi(X)$. The remaining cases are similar, only simpler. \square

The scattering calculus. As a rule, whenever the fibration $\varphi : \partial X \rightarrow Y$ is the identity map, i.e., whenever the fiber of φ is a point, we replace the identifier Φ with sc and talk about *scattering* instead of fibered cusp.

3. SCATTERING AND Φ -OPERATORS OF ORDER $-\infty$

We are going to show that for small $\varepsilon > 0$ there exists a non-natural injective algebra homomorphism from smoothing scattering operators on Y_ε to the algebra of smoothing fibered cusp-operators. This morphism Θ will be used in Section 4 to compute the Hochschild homology of the smoothing boundary ideal ${}^\Phi\mathcal{I}_\partial$.

The choice of a normal fibration near the boundary ∂X together with the boundary defining function ϱ_N yields an open neighborhood U_ε of the boundary and a diffeomorphism $U_\varepsilon \rightarrow \partial X \times [0, \varepsilon)$ for some $\varepsilon > 0$ such that $\varrho_N|_{U_\varepsilon}$ corresponds to the projection $\partial X \times [0, \varepsilon) \rightarrow [0, \varepsilon)$ onto the second factor. Moreover, $(U_\varepsilon)^2_\Phi := (\beta^2_\Phi)^{-1}(U_\varepsilon \times U_\varepsilon) \subseteq X^2_\Phi$ is an open neighborhood of the front face ff_Φ , and for $k \in \mathbb{C}$ we can define

$$\Psi^{-\infty, k}_{\Phi, \text{cpt}}(U_\varepsilon) := \varrho^{-k}_{\text{ff}_\Phi} \left\{ \kappa \in \mathcal{C}^\infty_c((U_\varepsilon)^2_\Phi, KD^{1/2}_\Phi((U_\varepsilon)^2_\Phi)) : \kappa \equiv 0 \text{ at } \partial(U_\varepsilon)^2_\Phi \setminus \text{ff}_\Phi \right\},$$

where $\varrho_{\text{ff}_\Phi} : (U_\varepsilon)^2_\Phi \rightarrow \overline{\mathbb{R}}_+$ is a defining function for the Φ -front face. Because of the compact support condition, $\Psi^{-\infty}_{\Phi, \text{cpt}}(U_\varepsilon)$ is closed under composition, and the canonical inclusion $\Psi^{-\infty}_{\Phi, \text{cpt}}(U_\varepsilon) \hookrightarrow \Psi^{-\infty}_\Phi(X)$ is a morphism of algebras.

Let $Y_\varepsilon := Y \times [0, \varepsilon)$. Then $U_\varepsilon \cong \partial X \times [0, \varepsilon) \xrightarrow{\varphi \times \text{id}} Y_\varepsilon$ is a locally trivial fiber bundle with fiber type F . As above, for $k \in \mathbb{C}$, let us define the compactly supported, smoothing scattering operators on Y_ε by

$$\Psi^{-\infty, k}_{sc, \text{cpt}}(Y_\varepsilon) := \varrho^{-k}_{\text{ff}_{sc}} \left\{ \kappa \in \mathcal{C}^\infty_c((Y_\varepsilon)^2_{sc}, KD^{1/2}_{sc}((Y_\varepsilon)^2_{sc})) : \kappa \equiv 0 \text{ at } \partial(Y_\varepsilon)^2_{sc} \setminus \text{ff}_{sc} \right\},$$

where $(Y_\varepsilon)^2_{sc}$ is the scattering double space, ff_{sc} the scattering front face of $(Y_\varepsilon)^2_{sc}$, and $\varrho_{\text{ff}_{sc}} : (Y_\varepsilon)^2_{sc} \rightarrow \overline{\mathbb{R}}_+$ a defining function for ff_{sc} . By the very definition, the fibration $U_\varepsilon \rightarrow Y_\varepsilon$ induces a fibration $\varphi^2_\Phi : (U_\varepsilon)^2_\Phi \rightarrow (Y_\varepsilon)^2_{sc}$ with fiber type $F \times F$.

Choose a fiber half-density $\nu \in \mathcal{C}^\infty(\partial X, \Omega^{1/2}_{\text{fiber}})$ such that $\int_{\varphi^{-1}(y)} \nu^2|_{\varphi^{-1}(y)} = 1$ for all $y \in Y$, and let

$$\begin{aligned} \nu^{(2)} &:= \nu \boxtimes \nu \in \mathcal{C}^\infty(\partial X \times \partial X, \Omega^{1/2}_{\text{fiber}}(\partial X) \boxtimes \Omega^{1/2}_{\text{fiber}}(\partial X)) \\ &= \mathcal{C}^\infty(\partial X \times \partial X, \Omega^{1/2}_{\text{fiber}}(\partial X \times \partial X)). \end{aligned}$$

We define

$$(13) \quad \Theta_\varepsilon : \Psi_{sc,cpt}^{-\infty,k}(Y_\varepsilon) \longrightarrow \Psi_{\Phi,cpt}^{-\infty,k}(U_\varepsilon) : \kappa \longmapsto (\varphi_\Phi^2)^*(\kappa) \otimes \nu^{(2)},$$

$$(14) \quad \Theta : \Psi_{sc,cpt}^{-\infty,k}(Y_\varepsilon) \xrightarrow{\Theta_\varepsilon} \Psi_{\Phi,cpt}^{-\infty,k}(U_\varepsilon) \hookrightarrow \Psi_\Phi^{-\infty,k}(X).$$

Before showing that the map Θ_ε is multiplicative, recall that the lifted Schwartz kernel κ_{AB} of the composition of two fibered cusp operators A and B can be computed using a triple version X_Φ^3 of the Φ -double space X_Φ^2 . This space is a compact manifold with corners that comes equipped with three b-fibrations

$$\pi_O^\Phi : X_\Phi^3 \rightarrow X_\Phi^2, \quad O \in \{C, F, S\},$$

where we have used the following projections:

$$\pi_C : X^3 \longrightarrow X^2 : (q_1, q_2, q_3) \longmapsto (q_1, q_3),$$

$$\pi_F : X^3 \longrightarrow X^2 : (q_1, q_2, q_3) \longmapsto (q_2, q_3),$$

$$\pi_S : X^3 \longrightarrow X^2 : (q_1, q_2, q_3) \longmapsto (q_1, q_2).$$

Using pull-back and push-forward under these b-fibrations, we obtain, up to a density factor,

$$(15) \quad \kappa_{AB} = (\pi_C^\Phi)_* \left((\pi_S^\Phi)^*(\kappa_A) \cdot (\pi_F^\Phi)^*(\kappa_B) \right).$$

The Φ -triple space X_Φ^3 has been constructed in [23], [42] in detail. Similarly, we can deal with the composition in the scattering calculus using the scattering triple space – for the triple space construction we refer to [28].

Let $(U_\varepsilon)_\Phi^3$ (resp. $(Y_\varepsilon)_{sc}^3$) be the fibered cusp (resp. scattering) triple space of U_ε (resp. Y_ε), and denote by $\pi_O^\Phi : (U_\varepsilon)_\Phi^3 \rightarrow (U_\varepsilon)_\Phi^2$ (resp. $\pi_O^{sc} : (Y_\varepsilon)_{sc}^3 \rightarrow (Y_\varepsilon)_{sc}^2$) the corresponding b-fibrations. Moreover, the fibration $U_\varepsilon \rightarrow Y_\varepsilon$ induces a fibration $\varphi_\Phi^3 : (U_\varepsilon)_\Phi^3 \rightarrow (Y_\varepsilon)_{sc}^3$ with fiber type F^3 , and the following diagram is commutative:

$$(16) \quad \begin{array}{ccccc} & & (U_\varepsilon)_\Phi^3 & & \\ & \swarrow \pi_F^\Phi & \downarrow \varphi_\Phi^3 & \searrow \pi_S^\Phi & \searrow \pi_C^\Phi \\ (U_\varepsilon)_\Phi^2 & & & (U_\varepsilon)_\Phi^2 & (U_\varepsilon)_\Phi^2 \\ \downarrow \varphi_\Phi^2 & & \downarrow \varphi_\Phi^2 & & \downarrow \varphi_\Phi^2 \\ & & (Y_\varepsilon)_{sc}^3 & & \\ & \swarrow \pi_F^{sc} & \downarrow \varphi_\Phi^2 & \searrow \pi_S^{sc} & \searrow \pi_C^{sc} \\ (Y_\varepsilon)_{sc}^2 & & & (Y_\varepsilon)_{sc}^2 & (Y_\varepsilon)_{sc}^2 \end{array}$$

Most important for us is the fact that the family of linear maps

$$\Theta_\varepsilon = \Theta_\varepsilon^{(k)} : \Psi_{sc,cpt}^{-\infty,k}(Y_\varepsilon) \rightarrow \Psi_{\Phi,cpt}^{-\infty,k}(U_\varepsilon), \quad k \in \mathbb{C},$$

is compatible with the composition of operators.

Theorem 3.1. *Let $A \in \Psi_{sc,cpt}^{-\infty,k}(Y_\varepsilon)$ and $B \in \Psi_{sc,cpt}^{-\infty,\ell}(Y_\varepsilon)$ be arbitrary. Then we have*

$$\Theta_\varepsilon^{(k+\ell)}(AB) = \Theta_\varepsilon^{(k)}(A)\Theta_\varepsilon^{(\ell)}(B) \in \Psi_{\Phi,cpt}^{-\infty,k+\ell}(U_\varepsilon).$$

Proof. It suffices to show that Θ_ε is multiplicative. We use the triple space construction and (15) for the kernel of the composition of two operators. Using the commutativity of (16), we obtain

$$\begin{aligned}\Theta_\varepsilon(A)\Theta_\varepsilon(B) &= \left[(\varphi_\Phi^2)^*\kappa_A \otimes \nu^{(2)}\right] \circ \left[(\varphi_\Phi^2)^*\kappa_B \otimes \nu^{(2)}\right] \\ &= (\pi_C^\Phi)_* \left(\left[(\pi_S^\Phi)^* \left((\varphi_\Phi^2)^*\kappa_A \otimes \nu^{(2)} \right) \right] \cdot \left[(\pi_F^\Phi)^* \left((\varphi_\Phi^2)^*\kappa_B \otimes \nu^{(2)} \right) \right] \right) \\ &= (\pi_C^\Phi)_* \left(\left[((\varphi_\Phi^3)^*(\pi_S^{sc})^*\kappa_A) \otimes \nu_S^{(2)} \right] \cdot \left[((\varphi_\Phi^3)^*(\pi_F^{sc})^*\kappa_B) \otimes \nu_F^{(2)} \right] \right) \\ &= (\varphi_\Phi^2)^* [(\pi_C^{sc})_* ((\pi_S^{sc})^*\kappa_A \cdot (\pi_F^{sc})^*\kappa_B)] \otimes \nu^{(2)} \\ &= (\varphi_\Phi^2)^*(\kappa_{AB}) \otimes \nu^{(2)} \\ &= \Theta_\varepsilon(AB).\end{aligned}$$

□

Recall the definitions

$$\begin{aligned}{}^{sc}\mathcal{I}_\partial &= \Psi_{sc,cpt}^{-\infty,\mathbb{Z}}(Y_\varepsilon)/\Psi_{sc,cpt}^{-\infty,-\infty}(Y_\varepsilon), \\ \Phi\mathcal{I}_\partial &= \Psi_\Phi^{-\infty,\mathbb{Z}}(X)/\Psi_\Phi^{-\infty,-\infty}(X).\end{aligned}$$

Note that the algebra ${}^{sc}\mathcal{I}_\partial$ does not depend on the particular choice of $\varepsilon > 0$. By (14), the family of linear maps $\Theta = \Theta^{(k)} : \Psi_{sc,cpt}^{-\infty,k}(Y_\varepsilon) \rightarrow \Psi_\Phi^{-\infty,k}(X)$ then induces the desired linear map $\Theta : {}^{sc}\mathcal{I}_\partial \rightarrow \Phi\mathcal{I}_\partial$.

Corollary 3.2. *The map $\Theta : {}^{sc}\mathcal{I}_\partial \rightarrow \Phi\mathcal{I}_\partial$ is a (non-natural) morphism of algebras.*

4. THE GENERALIZED HOCHSCHILD-KOSTANT-ROSENBERG MAP
AND THE HOMOLOGY OF THE SMOOTHING BOUNDARY IDEAL

We refer to Appendix B for a review of topological Hochschild homology and for a summary of the notation that we are going to use in the sequel.

The algebra map Θ constructed in Section 3 induces a map of complexes with the same name between the Hochschild complexes

$$\Theta : C_*({}^{sc}\mathcal{I}_\partial) \longrightarrow C_*(\Phi\mathcal{I}_\partial)$$

compatible with the boundary filtration. We will use this map to compare the two spectral sequences constructed with respect to this filtration. In fact, we get an isomorphism at ∂E^0 ; together with the convergence of the two spectral sequences, this implies that Θ induces an isomorphism on Hochschild homology.

The graded algebra ${}^{sc}\mathcal{I}_\partial^{[\mathbb{Z}]}$ is naturally isomorphic to the commutative algebra of Laurent polynomials in x with coefficients Schwartz functions on ΦN^*Y . Thus, the Hochschild-Kostant-Rosenberg map HKR induces an isomorphism from the Hochschild homology of ${}^{sc}\mathcal{I}_\partial^{[\mathbb{Z}]}$ to the space of forms (17). At the same time, Θ induces a map between the Hochschild complexes of ${}^{sc}\mathcal{I}_\partial^{[\mathbb{Z}]}$ and $\Phi\mathcal{I}_\partial^{[\mathbb{Z}]}$:

(17)

$C_k({}^{sc}\mathcal{I}_\partial^{[\mathbb{Z}]})$

$\downarrow \Theta$

$C_k(\Phi\mathcal{I}_\partial^{[\mathbb{Z}]})$

$\xrightarrow{HKR} \bigoplus_* \Lambda_S^*(\Phi N^*Y) \otimes \Lambda^{k-*}(\mathbb{C}[x, x^{-1}])$

$\nearrow K$

Our plan is to construct a map $K : C_k(\Phi \mathcal{I}_\partial^{[\mathbb{Z}]}) \rightarrow \bigoplus_* \Lambda_S^*(\Phi N^*Y) \otimes \Lambda^{k-*}(\mathbb{C}[x, x^{-1}])$ making the diagram (17) commutative and inducing an injection of $HH_*(\Phi \mathcal{I}_\partial^{[\mathbb{Z}]})$ into $\bigoplus_* \Lambda_S^*(\Phi N^*Y) \otimes \Lambda^{k-*}(\mathbb{C}[x, x^{-1}])$. The existence of K with the above properties will show that Θ becomes an isomorphism on $HH(\Phi \mathcal{I}_\partial^{[\mathbb{Z}]})$. In this section we will only use K on chains in $C_k(\Phi \mathcal{I}_\partial^{[\mathbb{Z}]})$, but in Section 6 it will be crucial to apply K on certain chains in the full graded boundary algebra $\Phi \mathcal{A}_\partial^{[\mathbb{Z}]}$; so we give here the general construction.

In order to define K we must first construct a covariant derivative on $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]}$. Let us explain this notion. Consider the sheaf of rings $\text{Pol}(\Phi N^*Y \times \mathbb{R})$ of polynomial functions on $\Phi N^*Y \times \mathbb{R}$. The algebra $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]}$ is a sheaf of $\text{Pol}(\Phi N^*Y \times \mathbb{R})$ -algebras over $\Phi N^*Y \times \mathbb{R}$, where the \mathbb{R} factor corresponds to the variable x . Consider the sheaf $\mathcal{V}^{\text{Pol}}(\Phi N^*Y \times \mathbb{R})$ of vector fields with polynomial coefficients on $\Phi N^*Y \times \mathbb{R}$. A covariant derivative on $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]}$ is then defined as a map of sheaves

$$\mathcal{V}^{\text{Pol}}(\Phi N^*Y \times \mathbb{R}) \otimes \Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]} \longrightarrow \Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]}$$

that is $\text{Pol}(\Phi N^*Y \times \mathbb{R})$ -linear in the first argument and a derivation in the second. More precisely,

$$(18) \quad \nabla_V(AB) = (\nabla_V A)B + A\nabla_V B,$$

$$(19) \quad \nabla_{fV} A = f\nabla_V A.$$

Fix a connection ∇ in the bundle $\Omega_{\text{fiber}}^{1/2}$ of fiberwise half-densities on ∂X and a connection in the fibration $\varphi : \partial X \rightarrow Y$, that is, a rule for lifting vectors from Y to ∂X . Pull back these connections through the maps $\partial X \times_Y \Phi N^*Y \rightarrow \partial X$ and $\Phi N^*Y \rightarrow Y$, respectively. They induce a connection in the fibration $\partial X \times_Y \Phi N^*Y \rightarrow \Phi N^*Y$ and a covariant derivative on the pull-back of $\Omega_{\text{fiber}}^{1/2}$ to $\partial X \times_Y \Phi N^*Y$. For a vector V tangent to ΦN^*Y , we denote by ∇_V the covariant derivative in the direction of the horizontal lift of V .

Lemma 4.1. *The covariant derivative ∇ maps the space*

$$\text{Pol}(\Phi N^*Y) \otimes S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2})$$

into the space

$$S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2}).$$

Proof. Clear in local coordinates in Y . □

The elements of the algebra $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [0]}$ are fiberwise half-densities on $\text{ff}_\Phi^{\text{int}}$. There are two ways of seeing $\text{ff}_\Phi^{\text{int}} \cong \partial X \times_Y \partial X \times_Y \Phi N^*Y$ as a fibration over $\partial X \times_Y \Phi N^*Y$, corresponding to the two projections onto the left factor (π_L) (resp. the right factor (π_R)). Thus, if $A \in \Phi \mathcal{A}_\partial^{\mathbb{Z}, [0]}$ and $s \in S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2})$, let

$$As := (\pi_L)_*(A\pi_R^*s) \in S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2}).$$

This defines a faithful action of $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [0]}$ on $S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2})$; hence, for $V \in \text{Pol}^\Phi N^*Y$, we can define the action of ∇_V on $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [0]}$ by duality:

$$\nabla_V(A)(s) := \nabla_V(A(s)) - A(\nabla_V(s)) \in S(\partial X \times_Y \Phi N^*Y, \Omega_{\text{fiber}}^{1/2}).$$

Finally, combining with the canonical action of $\partial/\partial x$ on $\mathbb{C}[x, x^{-1}]$, we get a covariant derivative on $\Phi \mathcal{A}_\partial^{\mathbb{Z}, [0]} \otimes \mathbb{C}[x, x^{-1}] = \Phi \mathcal{A}_\partial^{\mathbb{Z}, [\mathbb{Z}]}$, as claimed. Properties (18) and (19) are

easy to check. The covariant derivative preserves the ideal $\Phi\mathcal{I}_\partial^{\mathbb{Z},[\mathbb{Z}]}$. In the scattering case, the definition of ∇_V is independent of choices; so we write it simply as V . There exists a formula relating the algebra map Θ from Section 3 with ∇ .

Proposition 4.2. *Let $A_0, B_0 \in {}^s\mathcal{I}_\partial^{[0]} \otimes \mathbb{C}[x, x^{-1}] = {}^s\mathcal{I}_\partial^{[\mathbb{Z}]}$. Then*

$$(20) \qquad \qquad \qquad \nabla_V(\Theta(A_0))\Theta(B_0) = \Theta(V(A_0)B_0).$$

Proof. A little care should be exercised, because the “naive” statement $\nabla_V\Theta(A) = \Theta(V(A))$ does not hold. Essentially, the fiberwise density ν^2 used to define Θ is not parallel, while its volume (which was chosen to be 1) is. We can clearly reduce ourselves to proving the statement for $A_0, B_0 \in {}^s\mathcal{I}_\partial^{[0]}$. Note that ${}^s\mathcal{I}_\partial^{[0]}$ acts faithfully on $S(\Phi N^*Y)$ (by specializing from Φ to scattering the action described above). On the graded ideals, the map Θ can be described in terms of these actions by

$$\Theta(A_0)(s) = \pi^*(A_0(\pi_*(s \otimes \nu))) \otimes \nu,$$

where π is the projection $\partial X \times_Y \Phi N^*Y \rightarrow \Phi N^*Y$. Now $\pi_*(\pi^*(s_0) \otimes \nu^2) = s_0$ for all $s_0 \in S(\Phi N^*Y)$, because ν^2 has volume 1. Using this, the fact that $\pi_*(\nabla_V(\nu^2)) = V(\pi_*(\nu^2)) = V(1) = 0$, and the identity $\nabla_V(\pi^*s_0) = \pi^*(V(s_0))$, valid for any connection, the rest of the proof is straightforward. \square

We can now define the map K . This is done as in [35] along the lines of a map constructed in [33]. Let $A = A_0 \otimes \dots \otimes A_k$ be a k -Hochschild chain in $C_k(\Phi\mathcal{I}_\partial^{[\mathbb{Z}]})$. For V_1, \dots, V_k polynomial vector fields (or derivations in $\mathbb{C}[x, x^{-1}]$), define

$$K(A)(V_1, \dots, V_k) := \sum_{\sigma \in \Sigma_k} (-1)^{|\sigma|} \overline{\text{Tr}} \left(A_0 \nabla_{V_{\sigma(1)}} A_1 \dots \nabla_{V_{\sigma(k)}} A_k \right).$$

Here $\overline{\text{Tr}}$ denotes the fiberwise trace of smoothing suspended operators as explained in Section 2.

Remark 4.3. It is clear from the definition that K is still well-defined for tensor products $A = A_0 \otimes \dots \otimes A_k$ with $A_j \in \Psi_\Phi^{m_j, [\mathbb{Z}]}(X)$ such that $m_j \in \mathbb{C}$ and

$$\Re(m_0 + \dots + m_k) < -m = \dim(F),$$

because the fiberwise trace still makes sense. In that case $K(A)$ is a form with symbol coefficients of order $\Re(m_0 + \dots + m_k) + m$.

We will use this fact later on.

Proposition 4.4. (1) $K \circ b = 0$.

(2) $K \circ \Theta = HKR$.

(3) K is injective on $HH(\Phi\mathcal{I}_\partial^{[\mathbb{Z}]})$.

Proof. The first two claims are straightforward, using Proposition 4.2. For the third, we use the strategy of [35, Proposition 5.4.9] to show that every cycle is homologous to a cycle that belongs to the image of Θ . Then the first two assertions together with the Hochschild-Kostant-Rosenberg theorem yield the result. \square

The first two statements of Proposition 4.4 show that (17) commutes. From (17) it follows that Θ is injective on $HH({}^s\mathcal{I}_\partial^{[\mathbb{Z}]})$, because HKR is an isomorphism on homology. The third statement of Proposition 4.4 implies that Θ is also surjective from $HH({}^s\mathcal{I}_\partial^{[\mathbb{Z}]})$ to $HH(\Phi\mathcal{I}_\partial^{[\mathbb{Z}]})$. Note now that these homology groups are just

$\partial E^1({}^s\mathcal{I}_\partial)$, respectively $\partial E^1(\Phi\mathcal{I}_\partial)$. In conclusion Θ becomes an isomorphism at the level of ∂E^1 between the spectral sequences of ${}^s\mathcal{I}_\partial$ and $\Phi\mathcal{I}_\partial$.

The homology of ${}^s\mathcal{I}_\partial$ is a particular case of [14, Theorem 4.15] in the double-edge case. So we have

Proposition 4.5.

$$HH_k({}^s\mathcal{I}_\partial) = H_S^{2n+2-k}(\Phi N^*Y) \oplus H_S^{2n+1-k}(\Phi N^*Y) \frac{dx}{x}.$$

However, the proof from [14] is quite inexplicit, since it makes use of a Čech complex adapted to Hochschild homology. We can give a better proof of this result in the scattering case.

Form the spectral sequence ∂E with respect to the boundary filtration. We have just seen that

$$\partial E_{i,j}^1({}^s\mathcal{I}_\partial) = \Lambda_S^{i+j}(\Phi N^*Y)x^{-i} \oplus \Lambda_S^{i+j-1}(\Phi N^*Y)x^{-1-i}dx.$$

We claim that, up to sign,

$$(21) \quad d^1 = *_{sc} d *_{sc},$$

where d is the de Rham differential, and $*_{sc} := (\imath_{sc}^*)^{-1} * \imath_{sc}^*$ is the conjugate of Brylinski's symplectic duality operator by the dual \imath_{sc}^* of the scattering structure map $\imath_{sc} : {}^s\mathcal{TY}_\varepsilon \rightarrow TY_\varepsilon$. Indeed, this follows as in [3] from the formula of the normal operator of a commutator: if $A \in {}^s\mathcal{I}_\partial^i$, $B \in {}^s\mathcal{I}_\partial^j$, then $\mathcal{N}_\Phi^{[i+j-1]}([A, B]) = \{\mathcal{N}_\Phi^{[i]}(A), \mathcal{N}_\Phi^{[j]}(B)\}_{sc}$, where $\{ , \}_{sc} := (\imath_{sc}^*)^{-1} \{ , \} \imath_{sc}^*$ is the scattering Poisson bracket.

We now note that the isomorphism $*_{sc}$ splits according to x -degree as follows:

$$(22) \quad \begin{aligned} *_{sc} : \Lambda_S^k(\Phi N^*Y)x^l \oplus \Lambda_S^{k-1}(\Phi N^*Y)x^{l-1}dx \\ \longrightarrow \Lambda_S^{2n+2-k}(\Phi N^*Y)x^{l+k-n-1} \oplus \Lambda_S^{2n+1-k}(\Phi N^*Y)x^{l+k-n-2}dx. \end{aligned}$$

The homology of homogeneous forms in x is concentrated in homogeneity 0. Together with (21), it follows that

$$\partial E_{i,j}^2({}^s\mathcal{I}_\partial) = \begin{cases} H_S^{n+1-i}(\Phi N^*Y) \oplus H_S^{n-i}(\Phi N^*Y) \frac{dx}{x} & \text{if } j = n+1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\partial E(\Phi\mathcal{I}_\partial)$ degenerates at ∂E^2 . The convergence of these spectral sequences follows from Corollary B.9. Now a map between filtered complexes that induces an isomorphism on E^∞ is actually a quasi-isomorphism, provided the spectral sequences are convergent [24]. We summarize these remarks as follows.

Theorem 4.6. *The map Θ induces an isomorphism*

$$(23) \quad \begin{aligned} HH_k(\Phi\mathcal{I}_\partial) &\xleftarrow{\cong} HH_k({}^s\mathcal{I}_\partial) \cong H_S^{2n+2-k}(\Phi N^*Y) \oplus H_S^{2n+1-k}(\Phi N^*Y) \\ &\cong H^{n+1-k}(Y, \mathcal{O}(Y)) \oplus H^{n-k}(Y, \mathcal{O}(Y)), \end{aligned}$$

where $\mathcal{O}(Y) \rightarrow Y$ is the orientation bundle of Y . Moreover, $HH_k(\Phi\mathcal{I}_\partial)$ inherits the filtration by the total x -order, and

$$HH_k(\Phi\mathcal{I}_\partial)_j = \begin{cases} HH_k({}^s\mathcal{I}_\partial), & \text{if } j \geq k - n - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. For the last isomorphism in (23) we have used the Thom isomorphism. \square

5. THE HOMOLOGY OF THE SYMBOLIC Φ -ALGEBRAS

As Benameur and Nistor noticed [2], the homology of the algebra of symbols of a large class of pseudodifferential algebras can be computed using the spectral sequence argument of Wodzicki [43] and Brylinski and Getzler [3], [4]. In fact, this spectral sequence degenerates at E^2 . However, the convergence of this sequence is non-trivial because of the boundary filtration that plays a role in the definition of the Hochschild chain spaces.

For the algebra $\Psi^{\mathbb{Z}}(M)$, where M is a closed manifold, the convergence is almost tautological (see, however, Appendix B). It is also fairly obvious for the double-edge algebra [14], because of the vanishing of the double-edge Poisson bracket at the boundary. In the fibered cusp case, however, convergence is subtle. We prove in Appendix B a general convergence result which applies to this case, formalizing the corresponding result from [35].

The Hochschild chain spaces $C_k(\Phi\mathcal{A}_\sigma)$ admit a filtration given by the total operator order of the factors in the tensor product. This filtration is compatible with the boundary map; hence it induces a spectral sequence ${}^\sigma E$. The ${}^\sigma E^1$ term is just the homology of the graded algebra, which in turn is isomorphic to a space of exterior forms, the isomorphism being given by the Hochschild-Kostant-Rosenberg map:

$${}^\sigma E^1_{i,j}(\Phi\mathcal{A}_\sigma) = HH_{i+j}(\Psi^{[\mathbb{Z}],\mathbb{Z}}_\Phi(X))_{[i]} \simeq \Lambda^{i+j}_{[i]}(\Phi T^*X \setminus \{0\})[x^{-1}],$$

where as usual the subscript $[i]$ in $\Lambda^*_{[i]}$ stands for the space of homogeneous forms of degree i with respect to the natural \mathbb{R}_+ -action along the fibers of $\Phi T^*X \setminus \{0\}$. We claim that up to sign the boundary map d^1 equals $*_\Phi d *_\Phi$, where $*_\Phi = (\iota^*_\Phi)^{-1} * \iota^*_\Phi$, and $*$ is the symplectic duality operator on T^*X [3]. Indeed, this can be proved as in [4], using the form of the Poisson bracket (5). In fact, this claim also follows from the corresponding result in [4] for the interior of X , and then by continuity at the boundary.

Notice now that $*_\Phi$ splits according to homogeneity:

$$*_\Phi : \Lambda^k_{[l]}(\Phi T^*X \setminus \{0\}) \rightarrow \Lambda^{2N-k}_{[N-k+l]}(\Phi T^*X \setminus \{0\}),$$

and that the de Rham cohomology of homogeneous forms is concentrated in homogeneity 0. Indeed, on $\Lambda^*_{[p]}$, contraction with the radial vector field \mathcal{R} in the fibers of ΦT^*X is a homotopy between pI and 0. It follows that

(24)

$${}^\sigma E^2_{i,j}(\Phi\mathcal{A}_\sigma) = \begin{cases} H^{N-i}(\Phi S^*X \times S^1) \oplus H^{N-i-1}(\Phi S^*X|_{\partial X} \times S^1) & \text{if } j = N, \\ 0 & \text{otherwise.} \end{cases}$$

This proves in particular the degeneracy of the spectral sequence, because all the higher differentials d^k for $k > 1$ must vanish. We must prove convergence; that is, we must show that the ${}^\sigma E^\infty$ terms are isomorphic to the graded groups associated to $HH(\Phi\mathcal{A}_\sigma)$ with respect to the symbol filtration. This is unfortunately not obvious: the E^0 term of the spectral sequence has infinitely many nonzero components along the diagonals $\{(i,j) \in \mathbb{Z}^2; i+j = \text{constant}\}$. So standard diagram chasing will produce infinite sums of Hochschild chains. Such sums make sense (i.e., are asymptotically summable) only if they are uniformly bounded in the three filtrations from Appendix B. Moreover, as for any spectral sequence, we only get information about filtration quotients of the homology; we include in the definition

of convergence the condition that the “residual” part of the homology (i.e., those classes that admit representatives of arbitrarily low filtration order) vanishes.

We are going to prove the convergence of ${}^\sigma E(\Phi \mathcal{A}_\sigma)$ using Theorem B.4, by showing that assumptions H1–H6 hold. First, the fibered cusp symbol algebra satisfies the hypothesis of Lemma B.6; so H4 and H5 do hold. H6 is obvious. Condition H2 follows immediately from (24), likewise the first part of condition H1 (the existence of $\beta \in C_k^{i+1, \mathbb{Z}; \mathbb{Z}}$ such that $b(\beta) + \alpha \in C_{k-1}^{i-1, \mathbb{Z}; \mathbb{Z}}$).

For H3, we study the spectral sequence ${}^\sigma E$ (relative to the symbol filtration) of the graded algebra $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$ of $\Phi \mathcal{A}_\sigma$ with respect to the boundary filtration. Recall from (9) that $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$ is isomorphic to the algebra $\tilde{\Psi}_{\text{sus}(\Phi N^* Y) - \varphi}^{\mathbb{Z}}(\partial X) \otimes \mathbb{C}[x, x^{-1}]$.

Proposition 5.1. *The spectral sequence ${}^\sigma E(\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]})$ of the Hochschild complex of $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$ degenerates at ${}^\sigma E^2$.*

Proof. This result is essentially contained in [33]. Consider the short exact sequence of algebras

$$(25) \quad 0 \longrightarrow B_{\text{sus}} \longrightarrow \Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]} \longrightarrow F_{\text{sus}} \longrightarrow 0,$$

where B_{sus} is the ideal of those symbols in $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$ that vanish rapidly at the vertical sub-bundle of $\Phi T^* X$, and (25) serves as the definition of the algebra F_{sus} . This short exact sequence induces long exact sequences of ${}^\sigma E^1$ terms, but these actually decompose as short exact sequences of homogeneous differential forms

$$0 \longrightarrow {}^\sigma E_{ij}^1(B_{\text{sus}}) \longrightarrow {}^\sigma E_{ij}^1(\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}) \longrightarrow {}^\sigma E_{ij}^1(F_{\text{sus}}) \longrightarrow 0.$$

Thus we get long exact sequences both on the ${}^\sigma E^2$ terms (from the short exact sequences of ${}^\sigma E^1$ terms) and on Hochschild homology (induced from (25) by H -unitality of B_{sus}).

There exist many exterior derivations on the algebra $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$, namely the covariant derivatives ∇_V of Section 4, for V either equal to ∂_x or a vector field with polynomial coefficients on ΦNY . These derivations act on the Hochschild complex by contraction and by Lie derivative (see [20]). These actions descend to Hochschild homology and to the spectral sequences, and they commute with the differentials.

Since the polynomial functions f on ΦNY and in the variable x are central in $\Phi \mathcal{A}_\sigma^{\mathbb{Z}, [\mathbb{Z}]}$, we can define an “exterior product” action of such a function f on the Hochschild complex by

$$a_0 \otimes \dots \otimes a_k \xrightarrow{f \wedge} \sum_{i=0}^k (-1)^i a_0 \otimes \dots \otimes a_i \otimes f \otimes \dots \otimes a_k.$$

Again, this operation commutes with the Hochschild boundary map, and hence with all differentials in the spectral sequence. It is straightforward to check that, for a polynomial function f as above,

$$[e_{\nabla_V}, f \wedge] = V(f)I, \quad [L_{\nabla_V}, f \wedge] = V(f)\wedge, \quad f \wedge g \wedge = -g \wedge f \wedge,$$

where $V(f)$ is the V -derivative of the function f in the direction of the vector V . Also, on Hochschild homology, $e_{\nabla_U} e_{\nabla_V} = -e_{\nabla_V} e_{\nabla_U}$. This was shown by hand in [35], but it follows directly from the product structure of Hochschild cohomology $HH^*(A, A)$ and its action on homology (we are grateful to Colin Ingalls for this argument).

These algebraic properties of $\Phi\mathcal{A}_\sigma^{\mathbb{Z},[\mathbb{Z}]}$ and the explicit computation of $\sigma E^2(F_{\text{sus}})$ and $\sigma E^2(B_{\text{sus}})$ similar to [33] are enough to show that (i) the spectral sequences $\sigma E(F_{\text{sus}})$ and $\sigma E(B_{\text{sus}})$ degenerate at σE^2 , and (ii) the boundary map in the sequence of σE^2 terms vanishes. Inductively, we get a long exact sequence of σE^p terms that actually splits into short exact sequences, for $p \geq 1$. The Five Lemma shows that $\sigma E(\Phi\mathcal{A}_\sigma^{\mathbb{Z},[\mathbb{Z}]})$ also degenerates at σE^2 . \square

By Corollary B.8 we see that H3 is fulfilled. Now we use the notation of Appendix B to prove H1.

Proposition 5.2. *Let $a \in C_k^{i,j;l}(\Phi\mathcal{A}_\sigma)$ be a chain surviving at $\sigma E_{i,k-i}^2$, i.e., such that $d^0[a]_{\sigma E^0} = 0$ and $d^1[a]_{\sigma E^1} = 0$. If $k-i \neq N$ and $l \geq 1 + \max\{i + \delta_{i+1}^0, j + \delta_{j+1}^0\}$, then there exists $\beta \in C_k^{i+1,j+1;l}(\Phi\mathcal{A}_\sigma)$ such that $b(\beta) + a \in C_k^{i,j+1;l}$ and $b(\beta) + a$ is exact as a chain in $C_k^{[i],j+1;l}$.*

Proof. Let $[a]_{\sigma E^1}(\Phi\mathcal{A}_\sigma)$ be the form represented by a at $\sigma E_{i,k-i}^1(\Phi\mathcal{A}_\sigma)$. Then $*_\Phi[a]_{\sigma E^1}$ is closed, since a is a cycle, and it is exact because $k-i \neq N$. Moreover,

$$(k-i-N) *_\Phi[a]_{\sigma E^1} = d\iota_{\mathcal{R}} *_\Phi[a]_{\sigma E^1} = d *_\Phi \iota_\Phi^*(\alpha) \wedge [a]_{\sigma E^1},$$

where α is the canonical 1-form on T^*X . By Lemma B.7 the form $\iota_\Phi^*(\alpha) \wedge [a]_{\sigma E^1}$ can be represented by a chain $\beta_1 \in C_k^{[i+1],j+1;l}(\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}})$. \square

Thus H1 also holds for $\Phi\mathcal{A}_\sigma$; so we can apply Theorem B.4. An entirely similar analysis is done for the algebras $\Phi\mathcal{A}_{\partial,\sigma}$ and $\Phi\mathcal{I}_\sigma$. We summarize the results in the next theorem. The subscripts *rel* (respectively, *abs*) denote de Rham cohomology of forms vanishing to the boundary (respectively, smooth up to the boundary).

Theorem 5.3.

$$\begin{aligned} HH_k(\Phi\mathcal{I}_\sigma) &\cong H_{rel}^{2N-k}(\Phi S^*X \times S^1), \\ HH_k(\Phi\mathcal{A}_\sigma) &\cong H_{abs}^{2N-k}(\Phi S^*X \times S^1) \oplus H^{2N-1-k}(\Phi S^*X|_{\partial X} \times S^1), \\ HH_k(\Phi\mathcal{A}_{\partial,\sigma}) &\cong H^{2N-k}(\Phi S^*X|_{\partial X} \times S^1 \times S^1). \end{aligned}$$

Moreover, for $i \neq k-N$, $HH_k(\Phi\mathcal{A}_\sigma)_{[i]}$ and $HH_k(\Phi\mathcal{A}_{\partial,\sigma})_{[i]}$ vanish, and $HH_k(\Phi\mathcal{A}_\sigma)$ (respectively $HH_k(\Phi\mathcal{A}_{\partial,\sigma})$) can be represented by chains in $C_k^{k-N,k-n-1}(\Phi\mathcal{A}_\sigma)$ (resp. in $C_k^{k-N,k-n-1}(\Phi\mathcal{A}_{\partial,\sigma})$).

Proof. The isomorphisms follow from Theorem B.4. By analyzing the Φ -symplectic duality operator $*_\Phi$ (the homogeneities given in (22) are valid for absolute forms with symbol coefficients as well) and by Lemma B.7, we get the indices with respect to the boundary filtration as claimed. \square

We remark that the long exact sequence in Hochschild homology (which exists since $\Phi\mathcal{I}_\sigma$ is H -unital) coincides with the long exact sequence of σE^2 terms induced from the short exact sequence of σE^1 terms, which is just the de Rham absolute-relative cohomology long exact sequence.

6. THE BOUNDARY SEQUENCE

In order to compute $HH(\Phi\mathcal{A}_\partial)$, consider the short exact sequence of algebras

$$(26) \quad 0 \rightarrow \Phi\mathcal{I}_\partial \rightarrow \Phi\mathcal{A}_\partial \rightarrow \Phi\mathcal{A}_{\partial,\sigma} \rightarrow 0.$$

We have already computed the homologies of the two extremal algebras. By H -unitality of ${}^\Phi\mathcal{I}_\partial$ (an issue that we are not going to stress here), there exists a long exact sequence in Hochschild homology induced from (26). This long exact sequence is completely determined by the boundary maps

$$(27) \quad \delta_k : HH_k({}^\Phi\mathcal{A}_{\partial,\sigma}) \rightarrow HH_{k-1}({}^\Phi\mathcal{I}_\partial).$$

When looking at the associated graded algebras with respect to the boundary filtration, the result is quite simple, in the sense that all boundary maps vanish. We have already mentioned this fact in Proposition 5.1. This says basically (as shown in [33]) that there is no analog of the index map for suspended operators. It follows that the boundary maps δ decrease the boundary filtration order by at least 1. The only relevant dimensions are for $k \leq n+2$; otherwise $HH_{k-1}({}^\Phi\mathcal{I}_\partial) = 0$. Now note that $HH_{k+1}({}^\Phi\mathcal{A}_{\partial,\sigma})$ has representatives in boundary filtration $k-n$ (Theorem 5.3).

Let α be a homology class in $HH_{k+1}({}^\Phi\mathcal{A}_{\partial,\sigma})$. By H -unitality, there exists $A \in C_{k+1}({}^\Phi\mathcal{A}_{\partial,\sigma})$ representing the class α such that $b(A) \in C_k({}^\Phi\mathcal{I}_\partial)$. Moreover, we can assume that $A \in C_{k+1}^{N-k-1,k-n}({}^\Phi\mathcal{A}_{\partial,\sigma})$. From the fact that the boundary map in the associated graded algebras vanishes, we can also assume that $b(A) \in C_k^{k-n-1}({}^\Phi\mathcal{I}_\partial)$.

The boundary map δ is defined simply by $\delta(\alpha) := [b(A)]$. We do not know in general how the isomorphism from Theorem 4.6 associates a cohomology class to a cycle; however, we have been able to show that the boundary filtration order of the cycle $b(A)$ is $k-n-1$, which is exactly the filtration order where $HH_k({}^\Phi\mathcal{I}_\partial)$ is concentrated (Theorem 4.6). This means that $[b(A)]$ can be explicitly computed from its top component:

$$[b(A)] = {}_{*sc}K(b(A)_{[k-n-1]}) \in H_S^{2n+2-k}({}^\Phi NY \otimes \mathcal{L}).$$

Now use the Thom isomorphism between Schwartz-coefficient de Rham cohomology and cohomology of the base with coefficients in the orientation bundle:

$$[b(A)] = \int_{{}^\Phi NY/Y} {}_{*sc}K(b(A)_{[k-n-1]}) \in H^{n+1-k}(Y \otimes \mathcal{L}, \mathcal{O}(Y)).$$

Now let Q be a positive Φ -operator of order $(1,0)$ and let Q^z be its complex powers. The construction of complex powers in the Φ -setting is briefly discussed in Theorem 7.1. Notice the identity

$$Q^z b(A) = b(Q^z A) + z Q^z e_{\log Q}(A) + O(z^2),$$

where $e_{\log Q}(A)$ is the action of the exterior derivation $A_k \mapsto [\log Q, A_k]$ on the Hochschild complex given explicitly by (see also [20])

$$A_0 \otimes \dots \otimes A_k \mapsto [\log Q, A_k] A_0 \otimes \dots \otimes A_{k-1}.$$

We can then write

$$(28) \quad \begin{aligned} [b(A)] &= \int_{{}^\Phi NY/Y} {}_{*sc}K(Q^z b(A)_{[k-n-1]})|_{z=0} \\ &= \left[\int_{{}^\Phi NY/Y} {}_{*sc}K(b(Q^z A)_{[k-n-1]})|_{z=0} \right]_{[0]} \\ &\quad + \text{Res}_{z=0} \left[\int_{{}^\Phi NY/Y} {}_{*sc}K(Q^z e_{\log Q}(A)) \right]_{[0]}, \end{aligned}$$

where the subscript $[0]$ denotes the part of homogeneity 0 in x . We first examine the second term in (28). By Fubini's formula, it equals

$$-\frac{1}{(2\pi)^m} \left[\operatorname{Res}_{z=0} \int_{\Phi T^*X/Y} {}^*\phi \operatorname{HKR}(Q^z e_{\log Q}(A)) \right]_{[0]}.$$

We claim that this is just

$$\frac{1}{(2\pi)^m} \int_{\Phi S^*X/Y} \alpha.$$

Indeed, ${}^*\phi \operatorname{HKR}(Q^z e_{\log Q}(A))$ is a form with symbol coefficients on ${}^\Phi T^*X$ of homogeneity z . So the residue only depends on the principal symbol and is given by the above formula by a standard computation in Wodzicki-type formulas.

We claim now that the first term in (28) vanishes in cohomology, i.e., it is exact. This would be clear if $b(Q^z A)_{[k-n]}$ were equal to 0, since in that case the same computation that gives d^1 in $\partial E({}^\Phi \mathcal{I}_\partial)$ would give

$$(29) \quad {}^*_{sc} K(b(Q^z A)_{[k-n-1]}) = d {}^*_{sc} K(Q^z A)_{[k-n]}.$$

This holds for values of z of small enough real part. By the Hodge theorem, the space of exact forms is closed in the \mathcal{C}^∞ topology; hence the regularized value at $z = 0$ is also exact. In general, $b(Q^z A)_{[k-n]} \equiv 0$ modulo an entire family of chains of order $(N - k - 2 + z, [k - n])$ that vanishes at $z = 0$; hence the identity (29) is shown to be valid at $z = 0$ as in [35]. We summarize what we have obtained so far.

Proposition 6.1. *In terms of the identifications of Theorems 5.3 and 4.6, the boundary map (27) is given by*

$$\delta_k = \frac{1}{(2\pi)^m} \int_{\Phi S^*X/Y}.$$

The two copies of S^1 in Theorem 5.3 correspond to a concentrated way of writing the de Rham cohomology of homogeneous forms in the variables x and r . By Proposition 6.1 the boundary map sends $H^{2N-k-1}({}^\Phi S^*X|_{\partial X} \times S^1) \otimes \frac{dr}{r}$ to 0; hence (from the long exact sequence induced by (26)) this space lies in the image of the canonical map $HH_k({}^\Phi \mathcal{A}_\partial) \rightarrow HH_k({}^\Phi \mathcal{A}_{\partial, \sigma})$. The dimension of the sphere fibers of the fibration ${}^\Phi S^*X|_{\partial X} \rightarrow \partial X$ equals the dimension of the base, and so from the Leray spectral sequence it follows that ${}^\Phi S^*X|_{\partial X}$ is cohomologically a product:

$$H^*({}^\Phi S^*X|_{\partial X}) \simeq H^*(\partial X, \mathcal{O}(X)) \otimes H^*(S^{N-1}).$$

For $k \leq n + 2 \leq N$ we deduce that

$$H^{2N-k}({}^\Phi S^*X|_{\partial X} \times S^1) \simeq H^{N-k+1}(\partial X \times S^1, \mathcal{O}(X)).$$

We get from here a characterization of $HH({}^\Phi \mathcal{A}_\partial)$. Let I_{N-k+1} be the push-forward map (integral along the fiber)

$$I_{N-k+1} : H^{N-k+1}(\partial X \times S^1, \mathcal{O}(X)) \rightarrow H^{n-k+2}(Y \times S^1, \mathcal{O}(Y)).$$

Theorem 6.2.

$$HH_k({}^\Phi \mathcal{A}_\partial) \cong H^{2N-k-1}({}^\Phi S^*X|_{\partial X} \times S^1) \oplus \ker(I_{N-k+1}) \oplus \operatorname{coker}(I_{N-k}).$$

Proof. In the long exact sequence induced by (26) we write $HH_k({}^\Phi \mathcal{A}_\partial)$ as $\ker(\delta_k) \oplus \operatorname{coker}(\delta_{k+1})$. These spaces were identified above. \square

Corollary 6.3. *The dimension of $HH_0(\Phi\mathcal{A}_\partial)$ is 1, and so there exists a unique (up to a constant) continuous trace on $\Phi\mathcal{A}_\partial$.*

Proof. From Theorem 6.2 for $k = 0$ we get

$$\begin{aligned} H^{2N-k-1}(\Phi S^*X|_{\partial X} \times S^1) &= \mathbb{C}, \\ I_{N+1} &= 0, \\ \text{coker}(I_N) &= 0. \end{aligned}$$

Only the last equality needs some explanation. The integral map

$$I_{\partial X} : H^N(\partial X \times S^1, \mathcal{O}(X)) \rightarrow \mathbb{C}$$

is an isomorphism. The same is true for $I_Y : H^{n+1}(Y \times S^1, \mathcal{O}(Y)) \rightarrow \mathbb{C}$. But $I_{\partial X} = I_Y \circ I_N$; so I_N is also an isomorphism. \square

We remark that the boundary map and the homology of $\Phi\mathcal{A}_\partial$ depend strongly on the geometry of the boundary fibration.

7. TRACES AND GENERALIZED RESIDUES OF THE FIBERED CUSP CALCULUS

The computations of the previous sections show that $\dim_{\mathbb{C}} HH_0(A) = 1$ for $A = \Phi\mathcal{I}_\sigma, \Phi\mathcal{I}_\partial, \Phi\mathcal{A}_\sigma, \Phi\mathcal{A}_\partial, \Phi\mathcal{A}_{\partial,\sigma}$; thus, there exists, up to a multiplicative constant, a unique trace on these algebras. We are going to identify these traces as residues of the analytic continuation of a “double-zeta” function, and we are going to give, additionally, explicit formulas. The corresponding results about traces for the cusp calculus were obtained in [34, Section 5]. Some of the proofs depend on formal manipulations with zeta functions that can be found in [34]; so we will only state those results and leave the necessary changes of the proofs to the reader.

Theorem 7.1. *Let $Q \in \Psi_\Phi^{1,0}(X)$ be a positive Φ -operator, possibly with bundle coefficients. Then the family of complex powers $\mathbb{C} \ni z \mapsto Q^z$ is a holomorphic family of Φ -operators, $Q^z \in \Psi_\Phi^{z,0}(X)$.*

Proof. The complex powers of Q are constructed using the method of Bucicovschi [5], who extended the proof of Guillemin [12, Theorem 5.5] to the case of non-commutative symbols. We assume the reader to be familiar with these two papers for the purpose of this proof.

The algebra $\Psi_\Phi^m(X)$ has two symbol maps, which must be treated simultaneously. We first note that the complex powers of $\mathcal{N}_\Phi(Q)$ form a holomorphic family of suspended operators, $\mathcal{N}_\Phi(Q)^z \in \Psi_{\text{sus}(\Phi NY) - \varphi}^z(\partial X)$. This follows modulo $\Psi_{\text{sus}(\Phi NY) - \varphi}^{-\infty}(\partial X)$ from [5, Proposition 1.4], and is corrected to a true family of complex powers as in [12, Section 5]. It is clear that the principal (conormal) symbol of Q admits complex powers, and that

$$\sigma_z(\mathcal{N}_\Phi(Q)^z) = \sigma_1(Q)^z|_{\partial X}.$$

This allows us to start Bucicovschi’s induction argument: there exists a holomorphic family $Q_0(z)$ with $Q_0(0) = I$, $Q_0(1) = Q$, $\sigma_z(Q_0(z)) = \sigma_1(Q)^z$ and $\mathcal{N}_\Phi(Q_0(z)) = \mathcal{N}_\Phi(Q)^z$. This implies that

$$Q_0(z)Q_0(\tau)Q_0(z+\tau)^{-1} = 1 + R_1(z, \tau),$$

where $R_1(z, \tau) \in \Psi_\Phi^{-1,-1}(X)$ is holomorphic in the two variables. Notice that in [5] the error symbols R_j are sections in a von Neumann algebra bundle, whereas

here they have a component in the Fréchet algebra $\Psi^0_{\text{sus}(\Phi NY) - \varphi}(\partial X)$ (the boundary symbol). Nevertheless, there is basically no modification needed to prove [5, Propositions 1.3, 1.4] in our case; a new argument is only necessary to show that the map

$$\Psi^k_{\text{sus}(\Phi NY) - \varphi}(\partial X) \ni A \mapsto \int_0^1 Q^{-t} A Q^t dt \in \Psi^k_{\text{sus}(\Phi NY) - \varphi}(\partial X)$$

is surjective. This surjectivity is true modulo $\Psi^{k-1}_{\text{sus}(\Phi NY) - \varphi}(\partial X)$ (it reduces then to Bucicovschi's case of matrix-valued symbols), and also on $\Psi^{-\infty}_{\text{sus}(\Phi NY) - \varphi}(\partial X)$. Asymptotic completeness of $\Psi^0_{\text{sus}(\Phi NY) - \varphi}(\partial X)$ modulo the ideal $\Psi^{-\infty}_{\text{sus}(\Phi NY) - \varphi}(\partial X)$ ends the proof of the induction step.

By induction and asymptotic completeness of $\Psi_\Phi(X)$ modulo $\Psi_\Phi^{-\infty, -\infty}(X)$, we get an approximate complex powers family $Q(z)$ modulo $\Psi_\Phi^{-\infty, -\infty}(X)$, i.e.,

$$Q(z)Q(\tau)Q(z+\tau)^{-1} = 1 + R(z, \tau) \in 1 + \Psi_\Phi^{-\infty, -\infty}(X).$$

Again as in [12, Section 5] we show that $Q(z)$ differs from the true complex powers family Q^z by a holomorphic family in $\Psi_\Phi^{-\infty, -\infty}(X)$; so it follows that $Q^z \in \Psi_\Phi^{z, 0}(X)$ is a holomorphic family. \square

In particular, we see that the complex power Q^z is again a Φ -operator of complex order $z \in \mathbb{C}$. Extending the approach of Guillemin [12] and Bucicovschi [5], complex powers could recently be constructed simultaneously for a large class of boundary fibration structures and non-compact manifolds [1]; however, it is necessary to stress that in this generality complex powers are realized in the so-called Guillemin completion of the calculus, which contains slightly more smoothing operators than the original (small) calculus.

Proposition 7.2. *Let $\Omega \subseteq \mathbb{C}^2$ be open and connected with $[R, \infty) \times [R, \infty) \subseteq \Omega$ for some $R \in \mathbb{R}$, and let $A : \Omega \longrightarrow \Psi_\Phi^{m_0, k_0}(X)$ be holomorphic. Then the functions*

$$T_k : \Omega \cap \{\text{Re } z > k_0 + n + 1\} \cap \{\text{Re } \lambda > m_0 + N\} \longrightarrow \mathbb{C},$$

with $T_1 : (\lambda, z) \mapsto \text{Tr}(A(\lambda, z)\varrho_N^z Q^{-\lambda})$ and $T_2 : (\lambda, z) \mapsto \text{Tr}(A(\lambda, z)Q^{-\lambda}\varrho_N^z)$, are holomorphic. Moreover, there exist meromorphic functions $\tilde{T}_k : \Omega \longrightarrow \mathbb{C}$ with at most simple poles at $\lambda = m_0 + N - \ell$, $\ell \in \mathbb{N}_0$, and $z = k_0 + n + 1 - j$, $j \in \mathbb{N}_0$, which coincide with T_k on the component of Ω containing $[R', \infty) \times [R', \infty)$ for some $R' \in \mathbb{R}$.

For $A \in \Psi_\Phi^{m_0, k_0}(X)$, we let $Z_{\varrho_N, Q}(A)$ be the meromorphic extension of the holomorphic function

$$\{\text{Re } z > k_0 + n + 1\} \cap \{\text{Re } \lambda > m_0 + N\} \ni (\lambda, z) \longmapsto \text{Tr}(A\varrho_N^z Q^{-\lambda}),$$

which exists by Proposition 7.2; then $(\lambda, z) \mapsto z\lambda Z_{\varrho_N, Q}(A)(\lambda, z)$ is regular in a neighborhood of $0 \in \mathbb{C}^2$, and we can define $\text{Tr}_{\partial, \sigma}(A)$, $\widehat{\text{Tr}}_\partial(A)$ and $\widehat{\text{Tr}}_\sigma(A)$ by

$$(30) \quad \begin{aligned} z\lambda Z_{\varrho_N, Q}(A)(\lambda, z) = & \text{Tr}_{\partial, \sigma}(A) + \lambda \widehat{\text{Tr}}_\partial(A) + z \widehat{\text{Tr}}_\sigma(A) \\ & + \lambda^2 W(\lambda, z) + \lambda z W'(\lambda, z) + z^2 W''(\lambda, z), \end{aligned}$$

where W, W', W'' are holomorphic near $0 \in \mathbb{C}^2$ (and not unique!). Because of

$$\text{Tr}(A\varrho_N^z Q^{-\lambda} - A Q^{-\lambda} \varrho_N^z) = \text{Tr}(A(\text{id} - Q^{-\lambda} \varrho_N^z Q^\lambda \varrho_N^{-z}) \varrho_N^z Q^{-\lambda}),$$

we could have used in (30) also the zeta function $\tilde{Z}_{\varrho_N, Q}(A) : (\lambda, z) \mapsto \text{Tr}(A Q^{-\lambda} \varrho_N^z)$.

Furthermore, let ω_Φ be the canonical singular symplectic form on ${}^\Phi T^*X$, and let us denote by $i_{\mathcal{R}}$ (resp. $i_{x^2\partial_x}$) the contraction with the radial vector field \mathcal{R} in the fibers of ${}^\Phi T^*X$ (resp. with the canonical vector field $x^2\partial_x$, normal to the boundary). Recall that $\varrho_\sigma : {}^\Phi \overline{T}^*X \rightarrow \overline{\mathbb{R}}_+$ denotes a defining function for the Φ -cosphere bundle ${}^\Phi S^*X$. Moreover, the pull-back of the boundary defining function $\varrho_N : X \rightarrow \overline{\mathbb{R}}_+$ under the projection ${}^\Phi \overline{T}^*X \rightarrow \overline{\mathbb{R}}_+$ yields a defining function for the face ${}^\Phi \overline{T}^*X|_{\partial X}$, again denoted by ϱ_N for simplicity.

In the rest of this section, we summarize the properties of the functionals $\text{Tr}_{\partial,\sigma}$, $\widehat{\text{Tr}}_\partial$, and $\widehat{\text{Tr}}_\sigma$.

Proposition 7.3. *The double-residue $\text{Tr}_{\partial,\sigma} : \Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X) \rightarrow \mathbb{C}$ is a trace functional that is independent of the choice of Q . Moreover,*

$$\text{Tr}_{\partial,\sigma}(\Psi_{\Phi}^{-\infty,\mathbb{Z}}(X) + \Psi_{\Phi}^{\mathbb{Z},-\infty}(X)) = 0.$$

Thus, $\text{Tr}_{\partial,\sigma}$ defines a trace functional on ${}^\Phi \mathcal{A}_{\partial,\sigma}$. If $A \in {}^\Phi \mathcal{A}_{\partial,\sigma}$ is represented by $\sum_{k \leq k_0, m \leq m_0} A_{m,k}$ with $A_{m,k} \in \varrho_N^{-k} \varrho_\sigma^{-m} \mathcal{C}^\infty({}^\Phi S^*X|_{\partial X})$, then

$$\text{Tr}_{\partial,\sigma}(A) = (2\pi)^{-N} \int_{{}^\Phi S^*X|_{\partial X}} A_{-N, -(1+n)} i_{x^2\partial_x} i_{\mathcal{R}} \omega_\Phi^N.$$

The “unique” traces on ${}^\Phi \mathcal{A}_\sigma$ (resp. ${}^\Phi \mathcal{A}_\partial$) are then given by the composition of $\text{Tr}_{\partial,\sigma}$ with the natural projections ${}^\Phi \mathcal{A}_\sigma \rightarrow {}^\Phi \mathcal{A}_{\partial,\sigma}$ (resp. ${}^\Phi \mathcal{A}_\partial \rightarrow {}^\Phi \mathcal{A}_{\partial,\sigma}$).

Since $Z_{Q,x}(A)$ is entire for $A \in \Psi_{\Phi}^{-\infty,-\infty}(X)$, the linear functionals $\widehat{\text{Tr}}_\sigma$ and $\widehat{\text{Tr}}_\partial$ descend to linear functionals on $\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}/\Psi_{\Phi}^{-\infty,-\infty}$.

Proposition 7.4. *The restriction $\text{Tr}_\sigma : {}^\Phi \mathcal{I}_\sigma \rightarrow \mathbb{C}$ of $\widehat{\text{Tr}}_\sigma$ to ${}^\Phi \mathcal{I}_\sigma$ is a trace functional; it is independent of the choice of Q , and extends the Wodzicki residue for the double $2X$ of X by continuity from operators whose kernels are supported in the interior. If $A \in {}^\Phi \mathcal{I}_\sigma$ has the asymptotic expansion $\sum_{m \leq m_0} A_m$ with $A_m \in \varrho_N^\infty S^{[m]}({}^\Phi T^*X)$, i.e., $a_m := (\varrho_\sigma^m A_m)|_{{}^\Phi S^*X} \in \varrho_N^\infty \mathcal{C}^\infty({}^\Phi S^*X)$, then*

$$\text{Tr}_\sigma(A) = (2\pi)^{-N} \int_{{}^\Phi S^*X} a_{-N} i_{\mathcal{R}} \omega_\Phi^N.$$

To understand the functional $\widehat{\text{Tr}}_\sigma$, we need to extend the integral

$$(31) \quad \dot{\mathcal{C}}^\infty({}^\Phi S^*X) = \varrho_N^\infty \mathcal{C}^\infty({}^\Phi S^*X) \ni f \mapsto \int_{{}^\Phi S^*X} f i_{\mathcal{R}} \omega_\Phi^N$$

to functions $f \in \varrho_N^k \mathcal{C}^\infty({}^\Phi S^*X)$ with finite $k \in \mathbb{Z}$. For such an $f \in \varrho_N^k \mathcal{C}^\infty({}^\Phi S^*X)$, let

$$H_{\varrho_N}(f) : \{z \in \mathbb{C} : \text{Re } z > n + 1 - k\} \longrightarrow \mathbb{C} : z \mapsto \int_{{}^\Phi S^*X} \varrho_N^z f i_{\mathcal{R}} \omega_\Phi^N.$$

Then $H_{\varrho_N}(f)$ is holomorphic and admits a meromorphic extension $H_{\varrho_N}(f)$ to the complex plane with at most simple poles at $z = n + 1 - k - j$, $j \in \mathbb{N}_0$. Let $\oint_{{}^\Phi S^*X}^{\varrho_N} f i_{\mathcal{R}} \omega_\Phi^N$ be the regularized value of $H_{\varrho_N}(f)$ at $z = 0$. Note that $\oint_{{}^\Phi S^*X}^{\varrho_N}$ extends (31) as desired to $\varrho_N^k \mathcal{C}^\infty({}^\Phi S^*X)$, but depends mildly on the choice of the defining function ϱ_N .

Proposition 7.5. *Let $A \in \Psi_{\Phi}^{m_0,k_0}(X)$ be arbitrary, and suppose that the residue class $A + \Psi_{\Phi}^{-\infty,\mathbb{Z}}(X)$ corresponds to $\sum_{m \leq m_0} A_m$ with $A_m \in \varrho_N^{-k_0} S^{[m]}({}^\Phi T^*X)$, i.e.,*

$a_m := (\varrho_\sigma^m A_m)|_{\Phi S^* X} \in \varrho_N^{-k_0} \mathcal{C}^\infty(\Phi S^* X)$. Then

$$\widehat{\mathrm{Tr}}_\sigma(A) = \frac{1}{(2\pi)^N} \Phi\text{-}\int_{\Phi S^* X}^{\varrho_N} a_{-N} i_{\mathcal{R}} \omega_\Phi^N.$$

Proof. Consider $A\varrho_N^z$ for $\mathrm{Re}\, z > k_0 + n + 1$. Then the function

$$G_{\varrho_N} : \{z \in \mathbb{C} : \mathrm{Re}\, z > k_0 + n + 1\} \longrightarrow \mathbb{C} : z \mapsto \widehat{\mathrm{Tr}}_\sigma(A\varrho_N^z)$$

is holomorphic, admits a meromorphic extension G_{ϱ_N} to the complex plane with at most simple poles at $z = k_0 + n + 1 - j$, $j \in \mathbb{N}_0$, and from (30) we see that $\widehat{\mathrm{Tr}}_\sigma(A)$ coincides with the regularized value of G_{ϱ_N} at $z = 0$. Exactly as in Proposition 7.4 we obtain $G_{\varrho_N}(z) = (2\pi)^{-N} \int_{\Phi S^* X} b(z)_{-N} i_{\mathcal{R}} \omega_\Phi^N$, where the residue class of $A\varrho_N^z \in \Psi_\Phi^{m_0, k_0 - \mathrm{Re}\, z}(X)$ modulo $\Psi_\Phi^{-\infty, \mathbb{Z}}(X)$ is represented by the sum $\sum_{m \leq m_0} B(z)_m$ with $B(z)_m \in \varrho_N^{\mathrm{Re}\, z - k_0} S^{[m]}(\Phi T^* X)$ and $b(z)_m = (\varrho_\sigma^m B(z)_m)|_{\Phi S^* X}$. Since on the other hand we have $b(z)_m = \varrho_N^z a_m$, a further look at the definition of the Φ -integral $\Phi\text{-}\int^{\varrho_N}$ completes the proof. \square

Finally, for the functional $\widehat{\mathrm{Tr}}_\partial$ we need to consider an extension of the canonical trace $\overline{\mathrm{Tr}}$ on the space $\widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^{-\infty}(\partial X)$ to $\widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^M(\partial X)$ for arbitrary finite $M \in \mathbb{Z}$ (recall the canonical trace $\overline{\mathrm{Tr}}$ that was introduced in Section 2). To this end, fix a positive operator $Q_{\partial X} \in \widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^1(\partial X)$, and construct the holomorphic family $\mathbb{C} \ni \lambda \mapsto Q_{\partial X}^\lambda \in \widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^\lambda(\partial X)$ of complex powers. Note that we could take $Q_{\partial X} = \mathcal{N}_\Phi(Q)$ with $Q_{\partial X}^\lambda = \mathcal{N}_\Phi(Q)^\lambda = \mathcal{N}_\Phi(Q^\lambda)$ for $Q \in \Psi_\Phi^{1,0}(X)$, as above. Then, for $\widetilde{h} \in \widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^M(\partial X)$, the function $F_{Q_{\partial X}}(\widetilde{h}) : \lambda \mapsto \overline{\mathrm{Tr}}\left(\widetilde{h} Q_{\partial X}^{-\lambda}\right)$ is holomorphic in the domain $\{\lambda \in \mathbb{C} : \mathrm{Re}\, \lambda > M + N\}$, and admits a meromorphic extension to the whole plane with at most simple poles at $\lambda = M + N - j$, $j \in \mathbb{N}_0$. Let us denote by $\overline{\mathrm{Tr}}_{Q_{\partial X}}(\widetilde{h})$ the regularized value of $F_{Q_{\partial X}}(\widetilde{H})$ at $\lambda = 0$. By the very definition, $\overline{\mathrm{Tr}}_{Q_{\partial X}}$ is an extension of $\overline{\mathrm{Tr}}$ that, however, depends on the choice of the family $Q_{\partial X}$.

Proposition 7.6. *The restriction $\mathrm{Tr}_\partial : \Phi \mathcal{I}_\partial \rightarrow \mathbb{C}$ of $\widehat{\mathrm{Tr}}_\partial$ to $\Phi \mathcal{I}_\partial$ is a trace functional; it does not depend on the choice of Q . If $A \in \Phi \mathcal{I}_\partial$ corresponds to $\sum_{k \leq k_0} \varrho_{\mathrm{ff}\Phi}^{-k} \widetilde{A}_k$ with $\widetilde{A}_k \in \widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^{-\infty}(\partial X)$, then*

$$\mathrm{Tr}_\partial(A) = \overline{\mathrm{Tr}}(\widetilde{A}_{-(n+1)}).$$

For fibered cusp operators A of arbitrary order that vanish sufficiently fast at the front face, we also have a nice formula for $\widehat{\mathrm{Tr}}_\partial(A)$.

Proposition 7.7. *Let $A \in \Psi_\Phi^{m_0, -(n+1)}(X)$ be arbitrary, and suppose that the class $A + \Psi_\Phi^{\mathbb{Z}, -\infty}(X)$ is represented by $\sum_{k \leq -(n+1)} \varrho_{\mathrm{ff}\Phi}^{-k} \widetilde{A}_k$ with $\widetilde{A}_k \in \widetilde{\Psi}_{\mathrm{sus}(\Phi N^* Y) - \varphi}^{m_0}(\partial X)$. Then*

$$\widehat{\mathrm{Tr}}_\partial(A) = \overline{\mathrm{Tr}}_{Q_{\partial X}}(\widetilde{A}_{-(n+1)}),$$

where $Q_{\partial X} = \mathcal{N}_\Phi(Q)$.

Proof. Consider $AQ^{-\lambda}$ for $\lambda \in \mathbb{C}$ with $\mathrm{Re}\, \lambda > m_0 + N$. Then the function

$$H_Q : \{\lambda \in \mathbb{C} : \mathrm{Re}\, \lambda > m_0 + N\} \longrightarrow \mathbb{C} : \lambda \longmapsto \widehat{\mathrm{Tr}}_\partial(AQ^{-\lambda})$$

is holomorphic, admits a meromorphic extension H_Q to the complex plane with at most simple poles at $\lambda = m_0 + N - j$, $j \in \mathbb{N}_0$, and from (30) we see that $\widehat{\text{Tr}}_\partial(A)$ coincides with the regularized value of H_Q at $\lambda = 0$. Exactly as in Proposition 7.6 we see that we have $\widehat{\text{Tr}}_\partial(AQ^{-\lambda}) = \overline{\text{Tr}}(\widehat{B(\lambda)}_{-(n+1)})$ where the residue class of $AQ^{-\lambda} \in \Psi_\Phi^{m_0-\lambda, -(n+1)}(X)$ modulo $\Psi_\Phi^{\mathbb{Z}, -\infty}(X)$ corresponds to $\sum_{k \leq -(n+1)} e_{\text{ff}\Phi}^{-k} \widehat{B(\lambda)}_k$ with $\widehat{B(\lambda)}_k \in \widetilde{\Psi}_{\text{sus}(\Phi N^* Y) - \varphi}^{m_0-\lambda}(\partial X)$. Now, $\widehat{B(\lambda)}_{-(n+1)} = \widetilde{\mathcal{N}}_\Phi^{(-(n+1))}(A) \widetilde{\mathcal{N}}_\Phi^{(0)}(Q)^{-\lambda} = \widetilde{A}_{-(n+1)} Q_{\partial X}^{-\lambda}$, and a further look at the definition of $\overline{\text{Tr}}_{Q_{\partial X}}(\widetilde{A}_{-(n+1)})$ completes the proof. \square

Let us point out that the expression for $\widehat{\text{Tr}}_\partial(A)$ for $A \in \Psi_\Phi^{m,k}(X)$ with arbitrary $k \in \mathbb{Z}$ is more complicated.

8. COMPUTATION OF THE INDEX

Recall from [23] that a fibered cusp operator $A \in \Psi_\Phi^{p,k}(X)$ is Fredholm between appropriate weighted Sobolev spaces if and only if its Φ -principal symbol ${}^\Phi\sigma^{(p,k)}(A)$ and its normal symbol $\mathcal{N}_\Phi^{(k)}(A)$ are both invertible. Let A be such an operator. Then A is invertible modulo $\Psi_\Phi^{-1,-1}(X)$, and by a formal Neumann series argument there exists $B \in \Psi_\Phi^{-p,-k}(X)$ such that $P_1 := \text{id} - BA$ and $P_2 := \text{id} - AB$ both belong to $\Psi_\Phi^{-\infty,-\infty}(X)$, and, thus, are of trace class. With a little bit more effort and using a formula for the relative inverse of an operator with closed range which goes back at least as far as [41] (see also [8] and [11, Bemerkung 5.7]) we can actually find $B \in \Psi_\Phi^{-p,-k}(X)$ such that P_1 and P_2 are the orthogonal projections onto the kernel and cokernel of A .

Note that $A \otimes B$ then defines a Hochschild cycle in $HH_1(\Psi_\Phi^{\mathbb{Z},\mathbb{Z}}/\Psi_\Phi^{-\infty,-\infty})$. This simple but crucial observation was made by Melrose and Nistor [34], in the context of scalar cusp pseudodifferential operators. An analogous statement holds for any other algebra where elliptic operators admit pseudo-inverses. Melrose and Nistor extend the definition of this 1-cycle to operators acting on sections of a bundle \mathcal{E} . By Morita invariance, the homology of such an algebra is isomorphic to the homology of the scalar algebra. Then the boundary map of the long exact sequence of the pair $(\Psi_\Phi^{-\infty,-\infty}, \Psi_\Phi^{\mathbb{Z},\mathbb{Z}})$ applied to this cycle is a 0-Hochschild class of trace equal to the index of A .

A slight extension of this construction is carried out in [35] for a fully elliptic operator between sections of arbitrary vector bundles \mathcal{E}, \mathcal{F} . Replace the cycle $A \otimes B$ with $\text{Tr}(i_1 A P_2 \otimes i_2 B P_1)$, where P_1, P_2 are projections from a trivial bundle \mathbb{C}^q to \mathcal{E}, \mathcal{F} , i_1, i_2 are embeddings in \mathbb{C}^q , and Tr is the generalized trace functional that induces Morita invariance, $\text{Tr}((a_{ij}) \otimes (b_{ij})) = \sum a_{ij} \otimes b_{ji}$. Of course, Morita invariance does not apply in this general case, since operators from \mathcal{E} to \mathcal{F} do not form an algebra.

Now we formally follow the computation from [34] modified in [16] and applied in a similar situation in [14], and get the following result. The proof is included for the benefit of the reader.

Proposition 8.1. *The boundary map δ in the long exact sequence in Hochschild homology induced by the short exact sequence*

$$0 \longrightarrow \Psi_\Phi^{-\infty,-\infty} \longrightarrow \Psi_\Phi^{\mathbb{Z},\mathbb{Z}} \longrightarrow \Psi_\Phi^{\mathbb{Z},\mathbb{Z}}/\Psi_\Phi^{-\infty,-\infty} \longrightarrow 0$$

applied to a cycle $c \in C_1(\Psi_{\Phi}^{\mathbb{Z}, \mathbb{Z}}/\Psi_{\Phi}^{-\infty, -\infty})$ is given by

$$\mathbb{C} \ni \text{Tr}(\delta(c)) = \widehat{\text{Tr}}_{\partial}(f_{\log \varrho_N} c) - \widehat{\text{Tr}}_{\sigma}(e_{\log Q} c).$$

Here $f_{\log \varrho_N}$ and $e_{\log \varrho_N}$ are actions of the derivation $[\cdot, \log \varrho_N]$ on Hochschild 1-chains, namely

$$\begin{aligned} f_{\log \varrho_N}(A_0 \otimes A_1) &= A_0[A_1, \log \varrho_N], \\ e_{\log \varrho_N}(A_0 \otimes A_1) &= [A_0, \log \varrho_N]A_1. \end{aligned}$$

Proof. Let C be a lift of c to $C_1(\Psi_{\Phi}^{\mathbb{Z}, \mathbb{Z}})$. Then, from the definitions, we have $\delta(c) = [b(C)]_{HH_0(\Psi_{\Phi}^{-\infty, -\infty})}$, and hence

$$\text{Tr}(b(C)) = \text{Tr}(b(C)\varrho_N^z Q^{-\lambda})_{z=0, \lambda=0}.$$

To avoid introducing heavy algebraic notation, write formally $C = U \otimes V$. Then

$$\begin{aligned} \delta(c) &= \text{Tr}((UV - VU)\varrho_N^z Q^{-\lambda})_{z=0, \lambda=0} \\ &= \text{Tr} \left[(U(V - \varrho_N^z V \varrho_N^{-z})\varrho_N^z Q^{-\lambda}) + (Q^{\lambda} V Q^{-\lambda} - V)U\varrho_N^z Q^{-\lambda} \right]_{z=0, \lambda=0} \\ &= \text{Tr} \left[zU([V, \log \varrho_N] + zF_1(z))\varrho_N^z Q^{-\lambda} \right. \\ &\quad \left. + \lambda([\log Q, V] + \lambda F_2(\lambda))U\varrho_N^z Q^{-\lambda} \right]_{z=0, \lambda=0}, \end{aligned}$$

where $F_1(z), F_2(\lambda)$ are entire families of operators of fixed order. By Proposition 7.2, the result follows. \square

There exist more or less canonical choices for B and Q . More precisely, choose B to be an inverse of A up to $\Psi_{\Phi}^{-\infty, -\infty}$ and $Q = (x^k A^* A x^k)^{1/2p}$ modulo $\Psi_{\Phi}^{-\infty, -\infty}$. With these definitions, $\overline{AS}(A) := -\widehat{\text{Tr}}_{\sigma}([A, \log Q]B)$ and $\widehat{\text{Tr}}_{\partial}(A[B, \log \varrho_N])$ are independent of choices. Proposition 8.1 implies the following formula for the index of A :

Theorem 8.2.

$$\text{index}(A) = \overline{AS}(A) + \widehat{\text{Tr}}_{\partial}(A[B, \log \varrho_N]).$$

The local term can be identified with the coefficient of t^0 in the asymptotic expansion of the heat kernel of A^*A, AA^* (this is a universal local expression in the full symbol of A , and so we do not need to construct the heat kernel of Φ -operators in order to define it). We are able to determine the non-local term $\widehat{\text{Tr}}_{\partial}(A[B, \log \varrho_N])$ explicitly in geometric terms for certain differential operators on manifolds whose boundaries fiber over the circle S^1 .

Fix a connection in the boundary fibration, i.e., a rule for lifting horizontal vector fields. Fix a product Φ -metric on X , i.e., a smooth metric on the interior of X that with respect to local product coordinates close to the boundary looks like

$$(32) \qquad \frac{dx^2}{x^4} + \frac{d\sigma^2}{x^2} + g,$$

where σ is the variable in the circle and g is a family of metrics on the fibers of ϕ , independent of x . Let $E^{\pm} \rightarrow \partial X$ be vector bundles with fixed Hermitian metrics and compatible connections ∇ . Let $\mathcal{E}, \mathcal{F} \rightarrow X$ be vector bundles such that $\mathcal{E}|_{\partial X} \simeq \mathcal{F}|_{\partial X} \simeq E^+ \oplus E^-$. Let $A \in \text{Diff}_{\Phi}^1(X; \mathcal{E})$ be an elliptic Φ -differential operator of order 1 which near the boundary (for $x < \epsilon$) has the form

$$(33) \qquad A = \sigma(x) \left(x^2 \partial_x I_2 + \begin{bmatrix} -ix \tilde{\nabla}_{\partial_{\sigma}} & D^* \\ D & ix \tilde{\nabla}_{\partial_{\sigma}} \end{bmatrix} \right),$$

where D is a family of invertible operators on the fibers of the boundary, $\tilde{\nabla}_{\partial_\sigma} = \nabla_{\partial_\sigma} + \frac{1}{4} \text{Tr}(L_{\partial_\sigma} g)$, and $\sigma(x)$ is a linear isomorphism of $E^+ \oplus E^-$ that depends on x . This hypothesis is motivated by the fact that when the manifold X and the fibers of the boundary are spin manifolds so that the spin structures on X and on the fibers are compatible, a twisted Φ -Dirac operator on X with respect to the metric (32) has this form.

Theorem 8.3. *Under the above assumptions we have*

$$(34) \quad \text{index}(A) = \overline{AS}(A) - \frac{\lim_a \eta(\delta_x)}{2},$$

where the second term is the adiabatic limit of the eta invariant of the operator

$$\begin{bmatrix} -ix\tilde{\nabla}_{\partial_\sigma} & D^* \\ D & ix\tilde{\nabla}_{\partial_\sigma} \end{bmatrix}$$

on ∂X as $x \rightarrow 0$.

For twisted Dirac operators we identify the local term with the characteristic form $\hat{A}(X)\text{ch}(T)$, whose integral on X is convergent. A related formula has been obtained by Nye and Singer [38] for the Dirac operator on $X = S^1 \times \mathbb{R}^3$, where the boundary fibration is the projection $S^1 \times S^2 \rightarrow S^2$. The result announced above requires a sophisticated proof that will be given in [13].

APPENDIX A. SYMBOLIC AND SCHWARTZ SECTIONS OF VECTOR BUNDLES

For the convenience of the reader, we recall briefly the definitions of symbolic sections and Schwartz sections of certain vector bundles. An important notion for that is the radial compactification of a vector bundle [28].

It is easy to check that the *radial compactification map*

$$\text{RC} : \mathbb{R}^N \longrightarrow S_+^N := \{(\omega_0, \omega') \in S^N : \omega_0 \geq 0\} : z \longmapsto \langle z \rangle^{-1}(1, z)$$

with $\langle z \rangle := \sqrt{1 + |z|^2}$ identifies \mathbb{R}^N with the interior of the upper half sphere S_+^N , a compact manifold with boundary. The following basic observation is essential for the definitions of this section.

Proposition A.1. *Let $\varrho : S_+^N \rightarrow \overline{\mathbb{R}}_+$ be a boundary defining function. The radial compactification map RC induces via pull-pack isomorphisms $\dot{\mathcal{C}}^\infty(S_+^N) \rightarrow \mathcal{S}(\mathbb{R}^N)$ and $\varrho^{-m}\mathcal{C}^\infty(S_+^N) \rightarrow S^m(\mathbb{R}^N)$ where $\mathcal{S}(\mathbb{R}^N)$ is the space of Schwartz functions on \mathbb{R}^N , and $S^m(\mathbb{R}^N)$ the space of classical symbols of order $m \in \mathbb{C}$.*

Now let V be an N -dimensional vector space, and $\overline{V} := V \uplus (V \setminus \{0\})/\mathbb{R}_+$. We want to introduce a differentiable structure on \overline{V} . Choose a linear isomorphism $T : V \rightarrow \mathbb{R}^N$, and consider

$$\overline{T} : \overline{V} \longrightarrow S_+^N : \zeta \longmapsto \begin{cases} \text{RC}(Tv), & \zeta = v \in V, \\ \lim_{t \rightarrow \infty} \text{RC}(Ttv), & \zeta = [v] \in (V \setminus \{0\})/\mathbb{R}_+. \end{cases}$$

Then \overline{T} is a well-defined bijection that gives \overline{V} the structure of a differentiable manifold with boundary, diffeomorphic to S_+^N . Since each $A \in \text{GL}(\mathbb{R}^N)$ induces a

diffeomorphism $\Phi_A : S_+^N \rightarrow S_+^N$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^N & \xrightarrow{A} & \mathbb{R}^N \\ \text{RC} \downarrow & & \downarrow \text{RC} \\ S_+^N & \xrightarrow{\Phi_A} & S_+^N \end{array}$$

the differentiable structure on \overline{V} does not depend on the choice of the map T .

Everything obviously varies smoothly with parameters; thus, the same construction carries over to smooth vector bundles. Let X be a compact manifold with corners, and $\pi : V \rightarrow X$ a smooth vector bundle of rank N . Then there is a unique differentiable structure on $\overline{V} := V \uplus (V \setminus \{0\})/\mathbb{R}_+$ such that $\bar{\pi} : \overline{V} \rightarrow X$ becomes a smooth fiber bundle; for each $q \in X$, the fiber $\bar{\pi}^{-1}(q)$ is obtained by radially compactifying the fiber $\pi^{-1}(q)$. The bundle \overline{V} is called the *radial compactification* of V , whereas the boundary component $S(V) := (V \setminus \{0\})/\mathbb{R}_+$ of \overline{V} is said to be the *sphere bundle* of V .

Now let $E \rightarrow X$ be another vector bundle over X . Note that there are canonical embeddings $\text{RC} : V \rightarrow \overline{V}$ and $\text{RC}^* : \pi^*E \rightarrow \bar{\pi}^*E$ such that the following diagram commutes:

$$\begin{array}{ccc} \pi^*E & \xrightarrow{\text{RC}^*} & \bar{\pi}^*E \\ \text{pr} \downarrow & & \downarrow \text{pr} \\ V & \xrightarrow{\text{RC}} & \overline{V} \end{array}$$

Definition A.2. Let $\varrho_\sigma : \overline{V} \rightarrow \overline{\mathbb{R}}_+$ be a defining function for the boundary face $S(V)$ of \overline{V} . A section $s : V \rightarrow \pi^*E$ is said to be a *Schwartz* (resp. *classical*) *symbol of order* $m \in \mathbb{C}$ provided it is the pull-back under RC of a section $\bar{s} : \overline{V} \rightarrow \bar{\pi}^*E$ in $\bigcap_{m \in \mathbb{Z}} \varrho_\sigma^{-m} C^\infty(\overline{V}, \bar{\pi}^*E)$ (resp. in $\varrho_\sigma^{-m} C^\infty(\overline{V}, \bar{\pi}^*E)$). We write $S(V, \pi^*E)$ (resp. $S^m(V, \pi^*V)$) for the space of all Schwartz functions (resp. symbols) of order m . Moreover, in the paper we use the space $S^{\mathbb{Z}}(V, \pi^*E) := \bigcup_{m \in \mathbb{Z}} S^m(V, \pi^*E)$ of all symbols of integral order.

APPENDIX B. HOMOLOGY OF TOPOLOGICAL FILTERED ALGEBRAS

Let A be an algebra. The Hochschild complex of A (with coefficients in A) is defined by $C_k(A) := A^{\otimes(k+1)}$ with differential $d : C_k(A) \rightarrow C_{k-1}(A)$ given by

$$\begin{aligned} b(a_0 \otimes \dots \otimes a_k) &:= a_0 a_1 \otimes \dots \otimes a_k - a_0 \otimes a_1 a_2 \otimes \dots \otimes a_k + \dots \\ &\quad + (-1)^{k-1} a_0 \otimes \dots \otimes a_{k-1} a_k + (-1)^k a_k a_0 \otimes \dots \otimes a_{k-1}. \end{aligned}$$

If A is a topological algebra, it is more meaningful to involve the topology in the definition (see [6]). Thus, assuming that A is a nuclear Fréchet algebra, one replaces the algebraic tensor product with the projective (completed) tensor product. The most useful example is when A is the algebra of smooth functions on a compact manifold M . Then the classical Hochschild-Kostant-Rosenberg map

$$\begin{aligned} HKR : C_k(C^\infty(M)) &\rightarrow \Lambda^k(M), \\ a_0 \otimes \dots \otimes a_k &\mapsto a_0 d(a_1) \wedge \dots \wedge d(a_k), \end{aligned}$$

has the property $HKR \circ b = 0$ and becomes an isomorphism on the continuous Hochschild homology. This result extends easily to the case where A is the algebra

of smooth functions on a manifold with boundary or with corners, with Laurent behavior at some boundary hypersurfaces. In that case the continuous Hochschild homology is isomorphic via the *HKR* map to the space of exterior forms on the manifold, with Laurent behavior at the same hypersurfaces. The map *HKR* also computes the homology of the field of Laurent series in one variable, the ring of Laurent series in several variables and of polynomial rings; we implicitly assume these facts in the body of the paper.

Since the product in the topological algebra $\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X)$ is only separately continuous, we cannot directly apply the definition of continuous Hochschild homology as it is defined for instance in [6]. Instead, we use in a particular case the concept of Hochschild homology for topological filtered algebras formally introduced by Benaneur and Nistor [2] to cover exactly the case of pseudodifferential operators. The definition (35) of the chain spaces can be traced back to [4].

Let \mathcal{A} be a bi-filtered algebra with topology with two increasing filtrations $\{\mathcal{A}^{i,j}\}_{i,j \in \mathbb{Z}}$ adapted to the product structure. The model for \mathcal{A} is $\Psi_{\Phi}^{\mathbb{Z},\mathbb{Z}}(X)$. So we use the analogous notation for quotients, ideals and associated graded algebras. For example, $\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}}$ means the graded algebra of \mathcal{A} with respect to the first filtration.

Definition B.1. We say that \mathcal{A} is a *topological filtered algebra* if the following conditions are fulfilled:

- (1) $\mathcal{A} = \bigcup_{i,j \in \mathbb{Z}} \mathcal{A}^{i,j}$, and the topology on \mathcal{A} is the strict inductive limit topology (this implies that $\mathcal{A}^{i,j}$, $\mathcal{A}^{i,\infty}$ and $\mathcal{A}^{\infty,j}$ are closed in \mathcal{A}).
- (2) For each $i, j \in \mathbb{Z}$ the space $\mathcal{A}^{i,j}$ is a nuclear Fréchet space.
- (3) The multiplication map $\mathcal{A}^{i_1,j_1} \otimes \mathcal{A}^{i_2,j_2} \rightarrow \mathcal{A}^{i_1+i_2,j_1+j_2}$ is continuous with respect to the projective topology.
- (4) The canonical maps from $\mathcal{A}^{i,j}/\mathcal{A}^{-\infty,j}$ to the projective limit as $k \rightarrow -\infty$ of $\mathcal{A}^{i,j}/\mathcal{A}^{k,j}$, and from $\mathcal{A}^{i,j}/\mathcal{A}^{i,-\infty}$ to the projective limit as $k \rightarrow -\infty$ of $\mathcal{A}^{i,j}/\mathcal{A}^{i,k}$, are isomorphisms of topological vector spaces.
- (5) The canonical maps from $\mathcal{A}^{[i],j}/\mathcal{A}^{[i],-\infty}$ to the projective limit as $k \rightarrow -\infty$ of $\mathcal{A}^{[i],j}/\mathcal{A}^{[i],k}$, and from $\mathcal{A}^{i,[j]}/\mathcal{A}^{-\infty,[j]}$ to the projective limit as $k \rightarrow -\infty$ of $\mathcal{A}^{i,[j]}/\mathcal{A}^{k,[j]}$, are isomorphisms of topological vector spaces.

We emphasize that our definition is less general than the one considered in [2].

Since for fixed i, j the product map $\mathcal{A}^{j,j} \otimes \mathcal{A}^{j,j} \rightarrow \mathcal{A}$ is continuous, one defines the continuous Hochschild chain spaces as

$$(35) \quad C_k(\mathcal{A}) := \bigcup_{j \in \mathbb{Z}} \mathcal{A}^{j,j} \hat{\otimes}^{(k+1)}.$$

Our innovation consists in introducing three filtrations on $C_k(\mathcal{A})$, in the following way: a pure tensor $a = a_0 \otimes \dots \otimes a_k$ is said to belong to the linear space $c_k^{i,j;l}(\mathcal{A})$ if $a_0 \in \mathcal{A}^{i_0,j_0}, \dots, a_k \in \mathcal{A}^{i_k,j_k}$ with $i_0 + \dots + i_k \leq i$, $j_0 + \dots + j_k \leq j$, and such that for any subset $S \subset \{0, \dots, k\}$, $\sum_{r \in S} i_r \leq l$ and $\sum_{r \in S} j_r \leq l$. The space $C_k^{i,j;l}(\mathcal{A})$ is defined as the closure of $c_k^{i,j;l}(\mathcal{A})$ inside $C_k(\mathcal{A})$, endowed with the inductive limit topology. One has the obvious inclusions $\mathcal{A}^{j,j} \hat{\otimes}^{(k+1)} \subset C_k^{(k+1)j,(k+1)j;(k+1)j}(\mathcal{A})$ and $C_k^{i,j;l}(\mathcal{A}) \subset \mathcal{A}^{l,l} \hat{\otimes}^{(k+1)}$; thus

$$C_k(\mathcal{A}) = \bigcup_{i,j,l \in \mathbb{Z}} C_k^{i,j;l}(\mathcal{A}).$$

Since the map b is continuous, it extends to the topological chain spaces. The topological Hochschild homology of \mathcal{A} is then defined as the homology of the complex $(C_k(\mathcal{A}), b)$. The advantage of the tri-filtered spaces $C_k^{i,j;l}(\mathcal{A})$ is that

$$b(C_k^{i,j;l}(\mathcal{A})) \subset C_{k-1}^{i,j;l}(\mathcal{A}),$$

while no analogous property holds for $\mathcal{A}^{j,j} \hat{\otimes}^{(k+1)}$.

Since the filtrations are preserved by b , we can form spectral sequences for the homology of the Hochschild complex with respect to these filtrations. A fundamental difficulty appears when analyzing such spectral sequences: the E^0 terms live on a half-plane. Thus, tautological properties of first-quadrant spectral sequences (like collapsing or convergence) will need careful proofs. In particular, we will need to sum infinite series of chains.

From the definition of topological filtered algebras, every sum $\sum_{q=0}^{\infty} a_q$, $a_q \in \mathcal{A}^{k_1-q,k_2}$ is *asymptotically summable* in \mathcal{A}^{k_1,k_2} in the following sense: there exists $a \in \mathcal{A}^{k_1,k_2}$ such that for all $l \geq 0$ we have

$$a - \sum_{q=0}^l a_q \in \mathcal{A}^{k_1-l-1,k_2}.$$

The same property also holds for the second filtration. This property implies that every sequence $a_q \in C_k^{i-q,j;l}(\mathcal{A})$, $q \geq 0$, is asymptotically summable inside $C_k^{i,j;l}(\mathcal{A})$. We stress that this statement would fail if we let the third index l vary with q .

Now let $\mathcal{A}^{\mathbb{Z},\mathbb{Z}}$ be a topological filtered algebra such that $\bigcap_{k \in \mathbb{Z}} \mathcal{A}^{k,\mathbb{Z}} = 0$. Let ${}^\sigma E(\mathcal{A})$ denote the spectral sequence of the Hochschild complex of \mathcal{A} with respect to the first filtration. We say that ${}^\sigma E(\mathcal{A})$ degenerates at ${}^\sigma E^p$ if the higher differentials d^p, d^{p+1} , etc. all vanish. There exist natural filtrations on Hochschild homology induced from the filtrations of the Hochschild complex; namely, a homology class lives in the first filtration i if it can be represented by a cycle in $C_*^{i,\mathbb{Z};\mathbb{Z}}(\mathcal{A})$. Moreover, there exist natural edge maps

$$(36) \quad HH_{i+q}(\mathcal{A})_{[i]} \rightarrow {}^\sigma E_{i,q}^p(\mathcal{A})$$

from the graded group of $HH_{i+q}(\mathcal{A})$ with respect to the filtration into the spectral sequence. We say that ${}^\sigma E(\mathcal{A})$ converges (to the graded Hochschild homology) if it degenerates at ${}^\sigma E^p$ for some p , if the edge maps (to ${}^\sigma E^p$) are isomorphisms, and if moreover the “residual” homology $\bigcap_{i \in \mathbb{Z}} HH_k(\mathcal{A})_i$ vanishes.

Let $(C_*^{\mathbb{Z},\mathbb{Z};\mathbb{Z}}, d)$ be any tri-filtered homology complex, i.e., $C_* = \bigcup_{i,j,l \in \mathbb{Z}} C_*^{i,j;l}$. To increase readability of the next paragraphs we will use the following conventions: the subscript $*$ will be generally suppressed. We will work with chains in C , $C^{[\mathbb{Z}],\mathbb{Z};\mathbb{Z}}$, $C^{\mathbb{Z},[\mathbb{Z}];\mathbb{Z}}$ or $C^{[\mathbb{Z}],[\mathbb{Z}];\mathbb{Z}}$, and we will indicate the chain space we are working in by the filtration indices: for example, $\alpha \in C^{i,[j];l}$ denotes a chain in $C^{\mathbb{Z},[\mathbb{Z}];\mathbb{Z}}$. If $\alpha \in C^{i,j;l}$, we will denote again by α its image in the complex $C^{i,[j];l}$.

Lemma B.2. *For any spectral sequence associated to a \mathbb{Z} -filtered homology complex (C_*, d) and for all $p \in \mathbb{N}$, the edge maps $H_{i+q}(C)_{[i]} \rightarrow E_{i,q}^p$ are injective.*

Proof. The upper index in C_*^* denotes the filtration, which is assumed to be increasing. Let $[\alpha] \in H_{i+k}(C)$ be represented by a cycle $\alpha \in C_{i+k}^i$. Assume that the image of α in $E_{i,q}^p$ vanishes. Then there exists a class $[\beta_{i+p-1}] \in E_{i+p-1,i+q+1}^{p-1}$ such that $d^{p-1}[\beta_{i+p-1}] = [\alpha]_{E^{p-1}}$. This means that there exists a representative $\beta_{i+p-1} \in C^{i+p-1}$ of $[\beta_{i+p-1}]$ such that $\alpha - d(\beta_{i+p-1})$ represents the null

class at E^{p-1} . Inductively, construct $\beta_{i+p-2} \in C^{i+p-2}, \dots, \beta_i \in C^i$ such that $\alpha - d(\beta_{i+p-1} + \dots + \beta_i)$ represents the null class at σE^0 . But this means that $\alpha - d(\beta_{i+p-1} + \dots + \beta_i) \in C^{i-1}$, or in other words that $[\alpha] \in H_{i+q}(C)_{i-1}$, as claimed. \square

This lemma is proved using diagram chasing. The essential step in the proof is summing $\beta_{i+p-1} + \dots + \beta_i$ inside C^{i+p-1} . As long as the sum is finite, there is of course no problem. However, diagram chasing for other ingredients of convergence will involve summing infinite series of chains of descending filtration orders. This is in principle impossible, since the other filtrations might go up! So we must analyze each step of the diagram chasing very carefully.

Let $N \in \mathbb{Z}$ be fixed. We assume that for any $i, j \in \mathbb{Z}$, $j' \geq j$, $i' \geq i$ and $l \geq 1 + \max(i + \delta_0^i, j + \delta_j^0)$, the following assumptions hold for the complex C (these will be checked individually for the Hochschild chain complexes of the fibered cusp symbol algebras):

- H1. For any $k \neq i + N$ and any chain $\alpha \in C_k^{i,j;l}$ that is exact modulo $C^{i-1, \mathbb{Z}; \mathbb{Z}}$, there exists $\beta \in C^{i+1, j+1; l}$ such that $b(\beta) + \alpha \in C^{i, j+1; l}$ and $b(\beta) + \alpha$ is exact in $C^{[i], j+1; l}$.
- H2. Let $\alpha \in C_k^{i, j; l}$ be such that $b(\alpha) \in C^{i-2, j; l}$ and $k \neq i + N$. Then α is exact modulo $C^{i-1, \mathbb{Z}; \mathbb{Z}}$.
- H3. Let $\alpha \in C^{i, j; l} + C^{i', j-1; l}$ be such that α is exact in $C^{i, [j]; l}$ modulo $C^{i-1, [j]; l}$. Then there exists $\beta \in C^{i+1, j; l}$ such that $b(\beta) + \alpha \in C^{i, j; l} + C^{i', j-1; l}$ and $b(\beta) + \alpha$ is exact in $C^{[i], [j]; l}$.
- H4. Let $\alpha \in C^{i, j; l} + C^{i-1, j'; l}$ be such that α is exact in $C^{[i], j; l}$. Then there exists $\beta \in C^{i, j; l}$ such that $b(\beta) + \alpha \in C^{i-1, j'; l}$.
- H5. Let $\alpha \in C^{i, j; l} + C^{i', j-1; l} + C^{i-1, j'; l}$ be such that α is exact in $C^{[i], [j]; l}$. Then there exists $\beta \in C^{i, j; l}$ such that $b(\beta) + \alpha \in C^{i-1, j'; l} + C^{i', j-1; l}$.
- H6. $\bigcap_{i \in \mathbb{Z}} C^{i, j; l} = 0$, and any series $\sum_{i=i_0}^{-\infty} a_i$ with $a_i \in C_k^{i, j; l}$ for fixed j, k, l is asymptotically summable.

The assumptions H1 and H2 say slightly more than the fact that $\sigma E(C)$ degenerates at σE^2 and that $\sigma E_{i,*}^2 = 0$ for $* \neq N$. Assumption H3 is slightly stronger than the degeneracy of $\sigma E(C^{\mathbb{Z}, [\mathbb{Z}]; \mathbb{Z}})$ at σE^2 . Assumption H4 improves the degeneracy of $\partial E(\mathcal{A}^{[\mathbb{Z}], \mathbb{Z}})$ at ∂E^1 . Finally, H5 gives uniform bounds in the third filtration for exact chains. We remark here that this condition and a statement analogous to the second part of Lemma B.7 must be checked in the proof of the convergence result of [4].

Proposition B.3. *Assume that H1–H6 hold for the complex C . Let $\alpha \in C_k^{i, j; l}$ be an exact chain modulo $C^{i-1, \mathbb{Z}; \mathbb{Z}}$. If $k \neq i + N$, then there exists $\beta \in C^{i+1, j+1; l}$ such that $b(\beta) + \alpha \in C^{i-1, j; l}$. If $k = i + N$, then the same conclusion remains valid with possibly higher indices j, l .*

Proof. Assume first that $k \neq i + N$. By H1 there exists $\beta_0 \in C^{i+1, j+1; l}$ such that $b(\beta_0) + \alpha \in C^{i, j+1; l}$ and $b(\beta_0) + \alpha$ is exact in $C^{[i], j+1; l}$. By H4, choose $\beta'_1 \in C^{i, j+1; l}$ such that $b(\beta_0 + \beta'_1) + \alpha \in C^{i-1, j+1; l}$. Assume inductively that for $0 \leq p \leq q-1$ we have found $\beta_p \in C^{i-p+1, j+1; l}$ and $\beta'_q \in C^{i-q+1, j+1; l}$ such that

$$b(\beta_0 + \dots + \beta_{q-1} + \beta'_q) + \alpha \in C^{i-q, j+1; l} + C^{i-1, j; l}.$$

The initial step $q = 1$ was done above. Since α vanishes in $C^{i-q,[j+1];l}$, it follows that $b(\beta_0 + \dots + \beta_{q-1} + \beta'_q) + \alpha$ is exact in $C^{i-q,[j+1];l}$. By H3 there exists $\beta''_q \in C^{i-q+1,j+1;l}$ such that $b(\beta_0 + \dots + \beta_{q-1} + \beta'_q + \beta''_q) + \alpha \in C^{i-q,j+1;l} + C^{i-1,j;l}$ is exact in $C^{[i-q],[j+1];l}$. By H5 there exists $\beta'_{q+1} \in C^{i-q,j+1;l}$ such that

$$b(\beta_0 + \dots + \beta_{q-1} + \beta'_q + \beta''_q + \beta'_{q+1}) + \alpha \in C^{i-q-1,j+1;l} + C^{i-1,j;l}.$$

Setting $\beta_q := \beta'_q + \beta''_q$, we complete our induction argument.

By H6, the series $\sum \beta_q$ is asymptotically summable; so set $\beta := \sum_{q=0}^\infty \beta_q \in C^{i+1,j+1;l}_{k+1}$. Since

$$d(\beta) + \alpha \in C^{i-1,j;l} + \bigcap_{q>0} C^{i-q-1,j+1;l} = C^{i-1,j;l},$$

using H6, the conclusion follows for $k \neq i + N$. Assume now that $k = i + N$. The first step of the argument goes through if we allow for possibly higher \tilde{j}, \tilde{l} . In the rest of the proof we did not invoke H1 or H2. So everything goes through for these new indices \tilde{j}, \tilde{l} . □

There is something very nontrivial about this proposition. Even without caring about the third index (which is annoying but not serious), a simple application of H1 would yield a chain $\beta \in C^{i+1,j+1;l}$ such that $d(\beta) + \alpha \in C^{i-1,j+1;l}$. Then diagram chasing to write α as a boundary would lead to an infinite series of chains of decreasing order in i , but *increasing* in j . It is fortunate that we can reduce back the second index in the Φ -case (mainly thanks to the fact that H3 holds in this case), but we believe that for a multi-fibration pseudodifferential algebra this would not be possible.

Theorem B.4. *Let C be a tri-filtered complex satisfying assumptions H1–H6. Then the spectral sequence ${}^\sigma E(C)$ is convergent.*

Proof. We would like to invert the edge maps (36). By H1, H2 and H4 the ${}^\sigma E^2_{i,*}$ term is concentrated on the line $* = N$. Then let A be a class at ${}^\sigma E^2_{i,N}$. Then A can be represented by a chain $\alpha_1 \in C^{i,j;l}_{i+N}$ (note that in the case of the Hochschild complex associated to ${}^\Phi \mathcal{I}_\sigma$, j can be taken $j := i + N$, see Theorem 5.3). Surviving at ${}^\sigma E^2$ means that $b(\alpha_1) \in C^{i-1,j;l}_{i+N-1}$ and that $b(\alpha_1) \in C^{[i-1],j;l}_{i+N-1}$ is exact. Thus, by H4 there exists $\alpha_2 \in C^{i-1,j;l}_{i+N}$ such that $b(\alpha_2) + b(\alpha_1) \in C^{i-2,j;l}_{i+N-1}$. Since $(i+N-1) - (i-2) \neq N$ and $b(\alpha_2 + \alpha_1)$ is exact modulo anything, we can apply Proposition B.3 to get $\alpha_3 \in C^{i-1,j+1;l}_{i+N}$ such that $b(\alpha_1 + \alpha_2 + \alpha_3) \in C^{i-3,j;l}_{i+N-1}$. Inductively, using H4, construct $\alpha_q \in C^{i-q+2,j+1;l}_{i+N}$ such that $b(\alpha_1 + \dots + \alpha_q) \in C^{i-q,j;l}_{i+N-1}$. By H6, the sum $\alpha := \sum_{q=1}^\infty \alpha_q \in C^{i,j;l}_{i+N}$ is asymptotically summable and represents A at ${}^\sigma E^2$ by construction, and

$$b(\alpha) \in \bigcap_{q \in \mathbb{N}} C^{i-q,j;l}_{i+N-1} = 0.$$

So we have found a Hochschild cycle α that maps to A under the edge map (36), which means that the edge maps are isomorphisms (we have seen above that they are injective).

Now we prove the vanishing of the “residual homology” $\bigcap_{i \in \mathbb{Z}} HH_k(\mathcal{A})_i$. In fact, we prove directly that $HH_k(\mathcal{A})_{[k-N-1]} = 0$. Let $[\alpha]$ be a homology class represented by a cycle $\alpha \in C^{k-N-1,j;l}$. By H2 and Proposition B.3, there exists $\beta_1 \in C^{k-N,j+1;l}$

such that $b(\beta_1) + \alpha \in C^{k-N-2,j;l}$. Inductively, there exists $\beta_q \in C^{k-N-q+1,j+1;l}$ such that $b(\beta_1 + \dots + \beta_q) + \alpha \in C^{k-N-q-1,j;l}$. The sum $\beta := \sum_{q=1}^{\infty} \beta_q$ is summable in $C^{k-N,j+1;l}$, and moreover

$$b(\beta) + \alpha \in \bigcap_{q \in \mathbb{N}} C^{k-N-q-1,j;l} = 0,$$

which shows that α is exact. □

Remark B.5. Proposition B.3 remains valid if we allow any number of exceptional N 's in H1 and H2, with the obvious modifications. Thus, Theorem B.4 holds if we allow a *finite* number of such exceptional N 's. This is needed, for instance, for convergence in the case of the adiabatic algebra [35].

We are now going to establish some sufficient conditions for a topological filtered algebra and the associated symbolic spectral sequence to satisfy the conditions H3, H4, and H5. So for the rest of this section we assume that $\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}}$ is isomorphic as an algebra to $\mathbb{C}[\rho, \rho^{-1}] \otimes A$, where A is one of the algebras $\mathcal{L} \otimes \mathcal{A}^{[0],[0]}$ or $C^\infty(Y)[x^{-1}]$, for Y a compact manifold with boundary and x a boundary-defining function for ∂Y .

Lemma B.6. *Properties H4 and H5 are satisfied by \mathcal{A} under the above assumption.*

Proof. The algebra $\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}}$ is isomorphic to the tensor product $\mathcal{A}^{[0],[0]} \otimes \mathbb{C}[\rho, \rho^{-1}] \otimes \mathbb{C}[x, x^{-1}]$. By the Eilenberg-Zilber theorem, we can replace the Hochschild complex of $\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}}$ by the tensor product of the Hochschild complexes $C_*(\mathcal{A}^{[0],[0]}) \otimes C_*(\mathbb{C}[\rho, \rho^{-1}]) \otimes C_*(\mathbb{C}[x, x^{-1}])$, since the explicit quasi-isomorphisms between the two complexes (the Alexander-Whitney map and the shuffle product) preserve the two gradings and the third filtration index l . It is then enough to prove the statement for the algebra $\mathbb{C}[x, x^{-1}]$ itself. This is done by writing an explicit homotopy between the Hochschild complex $C_*(\mathbb{C}[x, x^{-1}])$ and the so-called small complex (see [20, Section 3.1]). Hence we get H5. Property H4 is proved in exactly the same way, by using formal Laurent series rather than Laurent polynomials in the variable x^{-1} . □

Note now that the algebra $\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}}$ and hence also $HH_*(\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}})$ have two gradings (depending on choices), given by the negative of the total degrees in ρ and x .

Lemma B.7. • *Let $[\alpha] \in {}^\sigma E_{i,k}^1(\mathcal{A}^{\mathbb{Z},[\mathbb{Z}]})$ be a class that corresponds to a class of degree (i, j) when seen as an element of $HH_{i+k}(\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}})$. Then there exists $\alpha \in C_{i+k}^{i,j;\max(i+\delta_0^i,j+\delta_j^0)}(\mathcal{A}^{\mathbb{Z},[\mathbb{Z}]})$ representing $[\alpha]$.*

• *Let $[\alpha] \in {}^\sigma E_{i,k}^1(\mathcal{A}^{\mathbb{Z},\mathbb{Z}})$ be a class that corresponds to a class of degree (i, j) when seen as an element of $HH_{i+k}(\mathcal{A}^{[\mathbb{Z}],\mathbb{Z}})$. Then there exists $\alpha \in C_{i+k}^{i,j;\max(i+\delta_0^i,j+\delta_j^0)}(\mathcal{A}^{\mathbb{Z},[\mathbb{Z}]})$ representing $[\alpha]$.*

Proof. It is enough to prove the statement for the algebras $C^\infty(Y)[x^{-1}]$, \mathcal{L} and $\mathbb{C}[\rho, \rho^{-1}]$ (which is obvious), and then to use the shuffle map to the Hochschild complex of the tensor product algebra. □

Corollary B.8. *Assume that the spectral sequence ${}^\sigma E(\mathcal{A}^{\mathbb{Z},[\mathbb{Z}]})$ degenerates at ${}^\sigma E^2$. Then condition H3 is fulfilled.*

Proof. Let $\alpha \in C^{i,j;l}$ be such that α is exact in $C^{i,j;l}$ modulo $C^{i-1,j;l}$. This means there exists $\beta_p \in C^{i+p,j;\mathbb{Z}}$ such that $b(\beta) + \alpha \in C^{i-1,j;\mathbb{Z}}$. Then α survives as the null class in ${}^\sigma E^p(\mathcal{A}^{\mathbb{Z},[\mathbb{Z}]})$, so by hypothesis also at ${}^\sigma E^2$. In other words, there exists $[\beta] \in {}^\sigma E^1$ with $d^1[\beta] = -[\alpha]$. Clearly $[\beta]$ is of (negative) degree $(i+1, j)$ in ρ and x . By Lemma B.7 we can choose a representative $\beta \in C^{i+1,j;l}$ provided $l \geq \max(i+1\delta_0^i, j+\delta_j^0)$. This is the desired chain β . \square

Corollary B.9. *Let \mathcal{A} be a topological filtered algebra such that the second filtration is constant. Assume that ${}^\sigma E(\mathcal{A})$ degenerates at ${}^\sigma E^2$, that $\mathcal{A}^{[\mathbb{Z}]} \simeq \mathbb{C}[\rho, \rho^{-1}] \otimes \mathcal{A}^{[0]}$, and that $\mathcal{A}^{[0]}$ is H -unital. Then ${}^\sigma E(\mathcal{A})$ is convergent.*

Proof. The issue is finding asymptotically summable chains in

$$C_k(\mathcal{A}) := \bigcup_{i,l \in \mathbb{Z}} C_k^{i,l}(\mathcal{A}).$$

Because the graded algebra is a tensor product, we see by Lemmas B.6 and B.7 that if $\alpha \in C^{i,l}$ is exact modulo $C^{i-1;\mathbb{Z}}$, then there exists $\beta \in C^{i+1,l}$ such that $b(\beta) + \alpha \in C^{i-1;l}$, provided $l \geq i+1+\delta_i^0$. This allows us to conclude that the edge maps are isomorphisms and that the residual homology vanishes, by constructing asymptotically summable series of chains via diagram chasing as in Theorem B.4. \square

The last result is implicitly used in [4].

REFERENCES

- [1] B. Ammann, R. Lauter, V. Nistor, and A. Vasy. *Complex powers and non-compact manifolds*, math.OA/0211305, preprint, November 2002.
- [2] M. Benaneur and V. Nistor, *Homology of complete symbols and noncommutative geometry*, Landsman, N. P. (ed.) et al., *Quantization of singular symplectic quotients*, Basel, Birkhäuser. Prog. Math. **198** (2001), 21–46.
- [3] J.-L. Brylinski, *A differential complex for Poisson manifolds*, J. Differential Geometry **28** (1988), 93–114. MR **89m**:58006
- [4] J.-L. Brylinski and E. Getzler, *The homology of algebras of pseudodifferential operators and the noncommutative residue*, K-Theory **1** (1987), 385–403. MR **89j**:58135
- [5] B. Bucicovschi, *An extension of the work of V. Guillemin on complex powers and zeta functions of elliptic pseudodifferential operators*, Proc. Amer. Math. Soc. **127** (1999), 3081–3090. MR **2000a**:58085
- [6] A. Connes, *Noncommutative Geometry*, Academic Press, New York - London (1994). MR **95j**:46063
- [7] A. Connes, *Gravity coupled with matter and the foundation of non-commutative geometry*, Comm. Math. Phys. **182** (1996), 155–176. MR **98f**:58024
- [8] H. O. Cordes, *On a class of C^* -algebras*, Math. Annalen **170** (1967), 283–313. MR **35**:749
- [9] Yu. V. Egorov and B.-W. Schulze, *Pseudodifferential operators, Singularities, Applications*, volume 93 of Operator Theory and Applications. Birkhäuser, Basel (1997). MR **98e**:35181
- [10] C. L. Epstein, R. B. Melrose, and G. A. Mendoza, *Resolvent of the Laplacian on strictly pseudoconvex domains*, Acta Math. **167** (1991), 1–106. MR **92i**:32016
- [11] B. Gramsch, *Relative Inversion in der Störungstheorie von Operatoren und Ψ -Algebren*, Math. Annalen **269** (1984), 27–71. MR **86j**:47065
- [12] V. Guillemin, *A new proof of Weyl's formula on the asymptotic distribution of eigenvalues*, Adv. in Math. **55** (1985), 131–160. MR **86i**:58135
- [13] R. Lauter and S. Moroianu. *An index formula on manifolds with fibered cusp ends*, preprint, 2002.
- [14] R. Lauter and S. Moroianu, *Homology of pseudo-differential operators on manifolds with fibered boundaries*, Journal Reine Angew. Math. **547** (2002), 207–234.

- [15] R. Lauter and S. Moroianu, *The index of cusp operators on manifolds with corners*, Ann. Global Anal. Geom. **21**, nr. 1 (2002), 31–49. MR **2003e**:58033
- [16] R. Lauter and S. Moroianu, *On the index formula of Melrose and Nistor*. Preprint Nr. 3, IMAR, Bucharest, March 2000.
- [17] R. Lauter and S. Moroianu, *Fredholm theory for degenerate pseudodifferential operators on manifolds with fibered boundaries*, Comm. Partial Differential Equations **26** (2001), 233–283. MR **2002e**:58052
- [18] R. Lauter and V. Nistor, *Analysis of geometric operators on open manifolds: a groupoid approach*, In N.P. Landsman, M. Pflaum, and M. Schlichenmaier, ed., *Quantization of Singular Symplectic Quotients*, vol. 198 of *Progress in Mathematics*, pp. 181–229. Birkhäuser, Basel - Boston - Berlin, 2001.
- [19] M. Lesch and M. J. Pflaum, *Traces on algebras of parameter dependent pseudodifferential operators and the eta-invariant*, Trans. Amer. Math. Soc. **352** (2000), 4911–4936. MR **2001b**:58042
- [20] J.-L. Loday, *Cyclic homology*, volume 301 of *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, Berlin - Heidelberg - New York (1992). MR **94a**:19004
- [21] F. Mantlik, *Norm closure and extension of the symbolic calculus for the cone algebra*, Ann. Global Anal. and Geometry **13** (1995), 339–376. MR **97a**:58183
- [22] R. R. Mazzeo, *Elliptic theory of differential edge operators I*, Comm. Partial Differential Equations **16** (1991), 1615–1664. MR **93d**:58152
- [23] R. R. Mazzeo and R. B. Melrose, *Pseudodifferential operators on manifolds with fibred boundaries*, Asian J. Math. **2** (1998), 833–866. MR **2000m**:58046
- [24] J. McCleary, *User's guide to spectral sequences*, volume 12 of *Mathematical Lecture Series*, Publish or Perish, Wilmington (1985). MR **87f**:55014
- [25] R. B. Melrose, *Analysis on manifolds with corners*, in preparation.
- [26] R. B. Melrose, *Pseudodifferential operators, corners and singular limits*, In *Proceeding of the International Congress of Mathematicians*, Kyoto, Springer-Verlag, Berlin - Heidelberg - New York (1990), 217–234. MR **93j**:58131
- [27] R. B. Melrose, *The Atiyah-Patodi-Singer index theorem*, volume 4 of *Research Notes in Mathematics*, A. K. Peters, Wellesley, Massachusetts (1993). MR **96g**:58180
- [28] R. B. Melrose, *Spectral and scattering theory for the Laplacian on asymptotically Euclidean space*, In M. Ikawa (ed.), *Spectral and Scattering Theory*, volume 162 of *Lecture Notes in Pure and Applied Mathematics*, pages 85–130, New York, 1994. Marcel Dekker Inc. *Proceedings of the Taniguchi International Workshop held in Sanda, November 1992*. MR **95k**:58168
- [29] R. B. Melrose, *The eta invariant and families of pseudodifferential operators*, Math. Res. Letters **2** (1995), 541–561. MR **96h**:58169
- [30] R. B. Melrose, *Geometric scattering theory*, Cambridge University Press, 1995. MR **96k**:35129
- [31] R. B. Melrose, *Fibrations, compactifications and algebras of pseudodifferential operators*, In L. Hörmander and A. Mellin, editors, *Partial Differential Equations and Mathematical Physics*, Birkhäuser, Boston, 1996, 246–261. MR **98j**:58117
- [32] R. B. Melrose, *Geometric optics and the bottom of the spectrum*, In F. Colombini and N. Lerner, editors, *Geometrical optics and related topics*, volume 32 of *Progress in nonlinear differential equations and their applications*, Birkhäuser, Basel - Boston - Berlin (1997).
- [33] R. B. Melrose and V. Nistor, *Higher index and η -invariants for suspended algebras of pseudodifferential operators*, unfinished manuscript.
- [34] R. B. Melrose and V. Nistor, *Homology of pseudodifferential operators I. Manifolds with boundary*, to appear in Amer. Math. J., Preprint, May 1996.
- [35] S. Moroianu, *Higher residues on the algebra of adiabatic pseudodifferential operators*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts (1999).
- [36] W. Müller, *Manifolds with cusps of rank one. Spectral theory and L^2 -index theorem*, *Lecture Notes in Mathematics* **1244**, Springer-Verlag, Berlin (1987). MR **89g**:58196
- [37] R. Nest and E. Schrohe, *Hochschild homology of Boutet de Monvel's algebra*, in preparation.
- [38] T. M. W. Nye and M. A. Singer, *An L^2 -index theorem for Dirac operators on $S^1 \times \mathbb{R}^3$* , J. Funct. Anal. **177** (2000), 203–218. MR **2002a**:58020
- [39] V. Nistor, *Groupoids and the integration of Lie algebroids*, J. Math. Soc. Japan **52** (2000), 847–868. MR **2002e**:58035

- [40] S. Rempel and B.-W. Schulze, *Complete Mellin and Green symbolic calculus in spaces with conormal asymptotics*, Ann. Global Anal. Geometry **4** (1986), 137–224. MR **89f**:58132
- [41] C. E. Rickart, *Banach algebras with an adjoint operation*, Ann. Math. **47** (1946), 528–550. MR **8**:159b
- [42] B. Vaillant, *Index- and spectral theory for manifolds with generalized fibered cusps*, Ph.D. thesis, University of Bonn (2001).
- [43] M. Wodzicki, *Cyclic homology of differential operators*, Duke Math. J. **54** (1987), 641–647. MR **88k**:32035
- [44] M. Wodzicki, *Noncommutative residue. I. Fundamentals*, In *K-theory, arithmetic and geometry* (Moscow, 1984–1986), Springer, Berlin (1987), 320–399. MR **90a**:58175
- [45] M. Wodzicki, *Cyclic homology of pseudodifferential operators and noncommutative Euler class*, C. R. Acad. Sci. Paris Sér. I Math. **306** (1988), 321–325. MR **89h**:58189
- [46] J. Wunsch, *Propagation of singularities and growth for Schrödinger operators*, Duke Math. J. **98** (1999), 137–186. MR **2000h**:58054

FACHBEREICH 17 - MATHEMATIK, UNIVERSITÄT MAINZ, D-55099 MAINZ, GERMANY
E-mail address: lauter@mathematik.uni-mainz.de

INSTITUTUL DE MATEMATICĂ AL ACADEMIEI ROMÂNE, P.O. Box 1-764, RO-70700 BUCHAREST, ROMANIA
E-mail address: moroianu@alum.mit.edu

MODERATE DEVIATION PRINCIPLES FOR TRAJECTORIES OF SUMS OF INDEPENDENT BANACH SPACE VALUED RANDOM VARIABLES

YIJUN HU AND TZONG-YOW LEE

ABSTRACT. Let $\{X_n\}$ be a sequence of i.i.d. random vectors with values in a separable Banach space. Moderate deviation principles for trajectories of sums of $\{X_n\}$ are proved, which generalize related results of Borovkov and Mogulskii (1980) and Deshayes and Picard (1979). As an application, functional laws of the iterated logarithm are given. The paper also contains concluding remarks, with examples, on extending results for partial sums to corresponding ones for trajectory setting.

1. INTRODUCTION AND MAIN RESULTS

Let $\{X_n\}$ be a sequence of i.i.d. \mathbf{R}^d -valued random variables, satisfying $EX_1 = 0$ and $\text{Var}(X_1) < +\infty$. Let \tilde{S}_n be the trajectories of sums of $\{X_n\}$, that is, $\tilde{S}_n(t) = \sum_{i=1}^{[nt]} X_i + (nt - [nt])X_{[nt]+1}$, $t \in [0, 1]$. Deshayes and Picard [19] have studied moderate deviations (MDs) for $\{\tilde{S}_n\}$ in $C[0, 1]$, which generalized corresponding results obtained by Borovkov [10] and Mogulskii [30]. Borovkov and Mogulskii [12] extended Deshayes and Picard's [19] results to independent, identically distributed (i.i.d.) random vectors $\{X_n\}$ with values in a complete locally convex Hausdorff topological vector space, under the crucial assumption that

$$\mu \stackrel{\Delta}{=} \mathcal{L}(X_1) \in \text{CLT}.$$

That is, the law $\mathcal{L}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right)$ converges weakly to a non-degenerate normal distribution. For more related results, see Borovkov and Mogulskii [12] and the references therein. However, the CLT for i.i.d. random vectors $\{X_n\}$ is not easily satisfied when $\{X_n\}$ take values in a general Banach space. Motivated by the observation above, in the present paper, we shall investigate the moderate deviation principle (MDP) for the trajectories, \tilde{S}_n , of sums of i.i.d. random vectors $\{X_n\}$ with values in a separable Banach space, and aim at removing the assumption employed by

Received by the editors March 28, 2001 and, in revised form, May 3, 2001.

2000 *Mathematics Subject Classification.* Primary 60F10.

Key words and phrases. Moderate deviations, trajectories, functional law of the iterated logarithm.

Supported in part by the National Natural Science Foundation of China and the Education Department of China.

Borovkov and Mogulskii [12], $\mu \triangleq \mathcal{L}(X_1) \in \text{CLT}$. As an application, the functional laws of the iterated logarithm are given.

Now, we turn to describing our main results in detail. Let \mathbf{N} be the set of positive integers. For $x \in \mathbf{R}$, $[x]$ denotes the greatest integer $k \leq x$. Throughout this paper, $(\mathbf{E}, \|\cdot\|)$ will denote a separable Banach space and \mathbf{E}^* its dual space. For $N \in \mathbf{N}$, we endow the product space \mathbf{E}^N with the product topology, which can be induced by the metric $d(\cdot, \cdot)$ given by $d(x, y) = \sum_{i=1}^N \|x_i - y_i\|$, $x = (x_1, \dots, x_N)$, $y = (y_1, \dots, y_N) \in \mathbf{E}^N$. Let $\{X_n\}$ be a sequence of independent \mathbf{E} -valued random vectors with common distribution μ such that $\mu \in WM_0^2$, that is, $Ef(X_1) = 0$ and $Ef^2(X_1) < \infty$ for every $f \in \mathbf{E}^*$. Let $(H, \|\cdot\|_H)$ be the reproducing kernel Hilbert space associated to μ (see Goodman, Kuelbs and Zinn [22]). A good example, which reveals the structure, is the Wiener measure μ on $\mathbf{E} = C[0, 1]$ with the supremum norm. Then the associated reproducing kernel Hilbert space is the so-called Cameron-Martin space. Let $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$. We denote by \tilde{S}_n the trajectories of sums of $\{X_n\}$. In other words,

$$\tilde{S}_n(t) = S_{[nt]} + (nt - [nt])X_{[nt]+1}, \quad t \in [0, 1].$$

Denote by $S_n(\cdot)$ the piecewise constant functions of sums of $\{X_n\}$, that is,

$$S_n(t) = \sum_{i=1}^{[nt]} X_i, \quad t \in [0, 1].$$

Let $C([0, 1], \mathbf{E})$ be the Banach space of all continuous mappings from $[0, 1]$ into \mathbf{E} equipped with the sup-norm, $\|\cdot\|_\infty$. Denote by $D([0, 1], \mathbf{E})$ the Banach space of all right-continuous mappings with left limits from $[0, 1]$ into \mathbf{E} equipped with the metric $d_\infty(f, g) = \sup_{0 \leq t \leq 1} \|f(t) - g(t)\|$. Given a set A , let A^c and \bar{A} stand for the complement and the closure of A , respectively.

Define a function $\Lambda : \mathbf{E} \rightarrow [0, +\infty]$

$$(1.1) \quad \Lambda(x) = \begin{cases} \frac{1}{2}\|x\|_H^2, & \text{if } x \in H, \\ +\infty, & \text{otherwise,} \end{cases}$$

and define a mapping $\tilde{\Lambda} : D([0, 1], \mathbf{E}) \rightarrow [0, +\infty]$

$$(1.2) \quad \tilde{\Lambda}(f) = \begin{cases} \int_0^1 \Lambda(g(s))ds, & f \in \mathcal{AC}, \\ +\infty, & \text{otherwise,} \end{cases}$$

where $\mathcal{AC} = \{f \in D([0, 1], \mathbf{E}); \text{ there exists } g \in L^1([0, 1], \mathbf{E}) \text{ such that } f(t) = \int_0^t g(s)ds \text{ for } t \in [0, 1]\}$, and $L^1([0, 1], \mathbf{E})$ is the space of \mathbf{E} -valued Bochner integrable functions on $[0, 1]$. It is readily seen that $\mathcal{AC} \subset C([0, 1], \mathbf{E})$.

Throughout this paper, for random vectors $\{Y_n\}$, we will write $Y_n \xrightarrow{P} 0$ if $Y_n \rightarrow 0$ in probability as $n \rightarrow \infty$.

Throughout this paper, let $\{b(n); n \geq 1\}$ be a positive sequence such that

$$(1.3) \quad \frac{b(n)}{\sqrt{n}} \rightarrow \infty, \quad \frac{b(n)}{n} \rightarrow 0.$$

The following conditions are assumed to be satisfied.

$$(1.4) \quad E \exp(\beta \|X_1\|) < \infty \text{ for some } \beta > 0.$$

$$(1.5) \quad \frac{S_n}{b(n)} \xrightarrow{P} 0.$$

It is well-known that under conditions (1.4) and (1.5), $\left\{\frac{S_n}{b(n)}\right\}$ in \mathbf{E} satisfies an MDP with speed $\left\{\frac{n}{b^2(n)}\right\}$ and the rate function Λ , defined by (1.1). By this we mean that, for every closed set $F \subset \mathbf{E}$,

$$(1.6) \quad \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{S_n}{b(n)} \in F \right\} \leq - \inf_{x \in F} \Lambda(x),$$

and for every open set $G \subset \mathbf{E}$,

$$(1.7) \quad \liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{S_n}{b(n)} \in G \right\} \geq - \inf_{x \in G} \Lambda(x).$$

Throughout this article, a rate function is understood to have compact level sets, i.e., $\{\Lambda \leq a\}$ is compact in \mathbf{E} for all $a \geq 0$. For this result of the partial sum see Goodman, Kuelbs and Zinn [22, Lemma 2.1 (V)], Chen [13, Theorem 2]; [15, Theorem 1], de Acosta [2], Ledoux [26] and Jiang [24].

It should be mentioned that recently Arcones [7] gave necessary and sufficient conditions for MDP for $\left\{\frac{S_n}{b(n)}\right\}$ in the real-valued case.

The main results of this paper are following.

Theorem 1.1. *Let $\pi : 0 = t_0 < t_1 < \cdots < t_N = 1$ be a partition of $[0, 1]$. Assume (1.4) and (1.5) hold, then $\left\{\frac{1}{b(n)}(S_n(t_1), \dots, S_n(t_N))\right\}$ in \mathbf{E}^N satisfies an MDP with speed $\left\{\frac{n}{b^2(n)}\right\}$ and a rate function \tilde{I}^π , that is, for every closed set F and open set G of \mathbf{E}^N ,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{1}{b(n)} (S_n(t_1), \dots, S_n(t_N)) \in F \right\} &\leq - \inf_{z \in F} \tilde{I}^\pi(z), \\ \liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{1}{b(n)} (S_n(t_1), \dots, S_n(t_N)) \in G \right\} &\geq - \inf_{z \in G} \tilde{I}^\pi(z), \end{aligned}$$

where $\tilde{I}^\pi : \mathbf{E}^N \rightarrow [0, +\infty]$ is given by

$$(1.8) \quad \tilde{I}^\pi(z) = \sum_{i=1}^N (t_i - t_{i-1}) \Lambda \left(\frac{z_i - z_{i-1}}{t_i - t_{i-1}} \right)$$

for $z = (z_1, \dots, z_N) \in \mathbf{E}^N$, $z_0 \triangleq 0$.

So does $\left\{\frac{1}{b(n)}(\tilde{S}_n(t_1), \dots, \tilde{S}_n(t_N))\right\}$ in \mathbf{E}^N with the same speed $\left\{\frac{n}{b^2(n)}\right\}$ and the same rate function \tilde{I}^π .

Remark 1.1. It is readily checked that \tilde{I}^π inherits the property of compact level sets from function Λ .

Theorem 1.2. *Assume (1.4) and (1.5) hold, then $\left\{\frac{\tilde{S}_n}{b(n)}\right\}$ in $C([0, 1], \mathbf{E})$ satisfies an MDP with speed $\left\{\frac{n}{b^2(n)}\right\}$ and rate function $\tilde{\Lambda}$ defined by (1.2), that is, for every closed set F and open set G of $C([0, 1], \mathbf{E})$,*

$$\limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n}{b(n)} \in F \right\} \leq - \inf_{f \in F} \tilde{\Lambda}(f),$$

$$\liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n}{b(n)} \in G \right\} \geq - \inf_{g \in G} \tilde{\Lambda}(g).$$

So does $\left\{ \frac{S_n(\cdot)}{b(n)} \right\}$ in $D([0, 1], \mathbf{E})$ with the same speed $\left\{ \frac{n}{b^2(n)} \right\}$ and the same rate function $\tilde{\Lambda}$.

Remark 1.2. From the proofs given in the next section, it can be seen that all lemmas, and thus Theorems 1.1 and 1.2 remain true if we substitute the underlying interval $[0, 1]$ with interval $[0, T]$ for each $T > 0$.

Remark 1.3. Theorem 1.2 has generalized the related results obtained by Borovkov and Mogulskii [12] and Deshayes and Picard [19].

If \mathbf{E} is of type p ($1 < p < 2$), then condition $E\|X_1\|^p < \infty$ implies $\frac{S_n}{n^{1/p}} \xrightarrow{P} 0$ (see Ledoux and Talagrand [27, pp. 190 and 259]). Therefore, an immediate corollary of Theorem 1.2 is following.

Corollary 1.1. *Suppose \mathbf{E} is of type p , $1 < p < 2$, and $E \exp(\beta\|X_1\|) < \infty$ for some $\beta > 0$. Then $\left\{ \frac{\tilde{S}_n}{n^{1/p}} \right\}$ satisfies an MDP in $C([0, 1], \mathbf{E})$ with speed $\left\{ \frac{n}{n^{2/p}} \right\}$ and rate function $\tilde{\Lambda}$ defined by (1.2).*

It should be mentioned that the large deviation principles (LDPs) for $\{\tilde{S}_n; n \geq 1\}$ have been studied extensively, see Varadhan [32], Borovkov [10], Mogulskii [30], Deshayes and Picard [19], Borovkov and Mogulskii [12], Schuette [31], Dembo and Zajic [16], Hu [23] and the references therein. LDPs for sample paths of vector-valued Lévy processes have also been proved by de Acosta [3]. Arcones [7] gave necessary and sufficient conditions for LDPs and MDPs for $\{S_n(\cdot); n \geq 1\}$ in the real-valued case.

The projective-limit method has been developed to successfully treat many problems. For a nice account of the theory, see, for example, Dembo and Zeitouni [18]. We do not see how to prove our result by the projective-limit method. A method of subsequences is devised to establish the MDP upper bound. The lower bound is proved by an interesting calculation which is rather explicit. Both upper and lower bounds are presented in Section 2. Our method is also different from that of Borovkov and Mogulskii [12]. It would be interesting to see our problem worked out along the line of the projective-limit method; comparison of the two approaches should be instructive. On the other hand, it should also be interesting to look on the MDPs for stochastic processes in general, see Arcones [7] for corresponding results in real-valued case in this direction.

The Functional Laws of the Iterated Logarithm will be considered in Section 3. In Section 4, some remarks will be given.

2. THE PROOFS OF MAIN RESULTS

We begin with several lemmas, which are important for proving the main results.

Lemma 2.1. *Given $\varepsilon > 0$, $0 < \delta \leq 1$, under condition (1.5), for all sufficiently large n , we have*

$$P \left\{ \max_{1 \leq k \leq [n\delta]+1} \|S_k\| \geq \varepsilon b(n) \right\} \leq 2P \left\{ \|S_{[n\delta]+1}\| \geq \frac{\varepsilon}{2} b(n) \right\},$$

$$P \left\{ \max_{1 \leq k \leq n} \|S_k\| \geq \varepsilon b(n) \right\} \leq 2P \left\{ \|S_n\| \geq \frac{\varepsilon}{2} b(n) \right\}.$$

Proof. We will prove only the first inequality, for the second can be treated similarly. By Ottaviani's inequality (see also Araujo and Gine, [6, pp. 110-111]), for every $A > 0$ and every integer $M \geq 1$,

$$P \left\{ \max_{1 \leq k \leq M+1} \|S_k\| \geq A \right\} \leq \frac{P \left\{ \|S_{M+1}\| \geq \frac{A}{2} \right\}}{1 - \max_{1 \leq k \leq M} P \left\{ \|S_{M+1-k}\| \geq \frac{A}{2} \right\}}.$$

Take $M = [n\delta]$, $A = \varepsilon b(n)$. By (1.5), we can steadily prove that for sufficiently large n ,

$$\max_{1 \leq k \leq [n\delta]} P \left\{ \|S_{[n\delta]+1-k}\| \geq \frac{\varepsilon}{2} b(n) \right\} \leq \frac{1}{2},$$

which implies the desired result. Lemma 2.1 is proved.

Lemma 2.2. *Under condition (1.4), $\left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \right\}$ and $\left\{ \frac{S_n(\cdot)}{b(n)} \right\}$ are exponentially equivalent in $D([0, 1], \mathbf{E})$, that is, for each $\delta > 0$,*

$$\limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ d_\infty \left(\frac{\tilde{S}_n(\cdot)}{b(n)}, \frac{S_n(\cdot)}{b(n)} \right) > \delta \right\} = -\infty.$$

Proof. Since $d_\infty(\tilde{S}_n(\cdot), S_n(\cdot)) \leq \max_{1 \leq k \leq n} \|X_k\|$, by Chebyshev's inequality, for each $\delta > 0$,

$$\begin{aligned} P \left\{ d_\infty \left(\frac{\tilde{S}_n(\cdot)}{b(n)}, \frac{S_n(\cdot)}{b(n)} \right) > \delta \right\} &\leq nP\{\|X_1\| > \delta b(n)\} \\ &\leq n \exp(-\beta \delta b(n)) E \exp(\beta \|X_1\|), \end{aligned}$$

where $\beta > 0$ is as in (1.4), from which the exponential equivalence of $\left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \right\}$ and $\left\{ \frac{S_n(\cdot)}{b(n)} \right\}$ follows. Lemma 2.2 is proved.

Following the proofs of Chen [13, Theorem 2]; [15, Theorem 1]), one can steadily prove the following version, Lemma 2.3, of the MDP for subsequences of i.i.d. sums.

Lemma 2.3. *Let $\{Y_j\}$ be a sequence of i.i.d. random variables with values in $(\mathbf{E}, \|\cdot\|)$, with common distribution $\mu \in WM_0^2$ and such that $E \exp(\alpha \|Y_1\|) < \infty$ for some $\alpha > 0$. Let $\{n_k\}$ and $\{a_k\}$ be positive integers and positive numbers, respectively, such that as $k \rightarrow \infty$,*

$$n_k \rightarrow \infty, \quad \frac{a_k}{\sqrt{n_k}} \rightarrow \infty, \quad \frac{a_k}{n_k} \rightarrow 0$$

and

$$\frac{S_{n_k}}{a_k} \xrightarrow{P} 0,$$

where $S_{n_k} = \sum_{j=1}^{n_k} Y_j$, $k \geq 1$.

Then $\left\{ \frac{S_{n_k}}{a_k} \right\}$ satisfies an MDP with speed $\left\{ \frac{n_k}{a_k^2} \right\}$ and the rate function $I(x) = \Lambda(x)$, where Λ is defined by (1.1).

Lemma 2.4. For $0 \leq s < t \leq 1$ fixed, let $W_n(s, t) = S_n(t) - S_n(s)$. Assume (1.4) and (1.5) hold, then $\left\{ \frac{W_n(s, t)}{b(n)} \right\}$ in \mathbf{E} satisfies an MDP with speed $\left\{ \frac{n}{b^2(n)} \right\}$ and a rate function $I_{s, t}$, that is, for every closed set F and open set G of \mathbf{E} ,

$$\limsup_{k \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{W_n(s, t)}{b(n)} \in F \right\} \leq - \inf_{x \in F} I_{s, t}(x),$$
$$\liminf_{k \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{W_n(s, t)}{b(n)} \in G \right\} \geq - \inf_{x \in G} I_{s, t}(x),$$

where $I_{s, t} : \mathbf{E} \rightarrow [0, \infty]$ is defined as follows

(2.1)
$$I_{s, t}(x) = (t - s) \Lambda \left(\frac{x}{t - s} \right) = \frac{1}{t - s} \Lambda(x), \quad x \in \mathbf{E}.$$

Proof of Lemma 2.4. The function $I_{s, t}$ defined by (2.1) has compact level sets because Λ does, see Goodman, Kuelbs and Zinn [22, Lemma 2.1(V)] and Chen [13, Theorem 2]; [15, Theorem 1].

Given $0 \leq s < t \leq 1$, define $n_k = [kt] - [ks]$ and $a_k = b(k)$, $k \geq 1$. Then $W_k(s, t)$ has the same distribution as S_{n_k} and hence $\frac{W_k(s, t)}{b(k)}$ has the same distribution as $\frac{S_{n_k}}{a_k}$. From (1.5) and Lemma 2.1 it follows that $\frac{S_{n_k}}{a_k} \xrightarrow{P} 0$. Taking into account the fact that $\frac{k}{b^2(k)} \cdot \frac{a_k^2}{n_k} \rightarrow \frac{1}{t - s}$ as $k \rightarrow \infty$, Lemma 2.4 follows from Lemma 2.3. Lemma 2.4 is proved.

Proof of Theorem 1.1. Given a partition, π , of $[0, 1]$, $\pi : 0 = t_0 < t_1 < \cdots < t_N = 1$, denote $W_n^\pi = (S_n(t_1), S_n(t_2) - S_n(t_1), \cdots, S_n(t_N) - S_n(t_{N-1}))$. Then by Lemma 2.4, Lynch and Sethuraman [28, Corollary 2.9] (see also Dembo and Zeitouni [18, Ex. 4.2.7]), $\left\{ \frac{W_n^\pi}{b(n)} \right\}$ satisfies an MDP in \mathbf{E}^N with speed $\left\{ \frac{n}{b^2(n)} \right\}$ and rate function I^π given by

$$I^\pi(z) = \sum_{i=1}^N (t_i - t_{i-1}) \Lambda \left(\frac{z_i}{t_i - t_{i-1}} \right)$$

for $z = (z_1, \cdots, z_N) \in \mathbf{E}^N$.

Applying the contraction principle (see Dembo and Zeitouni [18, Theorem 4.2.1]) to the continuous one-to-one map $(z_1, \cdots, z_N) \rightarrow (z_1, z_1 + z_2, \cdots, \sum_{i=1}^N z_i)$ on \mathbf{E}^N , from the MDP for the sequence $\left\{ \frac{W_n^\pi}{b(n)} \right\}$ it follows that $\left\{ \frac{1}{b(n)} (S_n(t_1), \cdots, S_n(t_N)) \right\}$ satisfies an MDP in \mathbf{E}^N with speed $\left\{ \frac{n}{b^2(n)} \right\}$ and rate function defined by the equality (1.8). Lemma 2.2 implies that the sequences $\left\{ \frac{1}{b(n)} (S_n(t_1), \cdots, S_n(t_N)) \right\}$ and $\left\{ \frac{1}{b(n)} \left(\tilde{S}_n(t_1), \cdots, \tilde{S}_n(t_N) \right) \right\}$ in \mathbf{E}^N are exponentially equivalent. Hence, by Dembo and Zeitouni [18, Theorem 4.2.13], $\left\{ \frac{1}{b(n)} \left(\tilde{S}_n(t_1), \cdots, \tilde{S}_n(t_N) \right) \right\}$ in \mathbf{E}^N satisfies the same MDP as that for $\left\{ \frac{1}{b(n)} (S_n(t_1), \cdots, S_n(t_N)) \right\}$. Theorem 1.1 is proved.

Proof of Theorem 1.2. The compactness of level sets of $\tilde{\Lambda}$ will be proved in Appendix B.

Upper bound. We first show that for every $f \in C([0, 1], \mathbf{E})$, and every $\epsilon > 0$, there exists a ball of f , $B(f, \rho) = \{g; \|f - g\|_\infty < \rho\}$ for some $\rho = \rho(f, \epsilon) > 0$, such that

$$(2.2) \quad \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in B(f, \rho) \right\} \leq -\tilde{\Lambda}(f) + \epsilon.$$

To this end, given $f \in C([0, 1], \mathbf{E})$ and $\epsilon > 0$, we first consider the case where $\tilde{\Lambda}(f) < \infty$.

For every partition, π , of $[0, 1]$, $\pi: 0 = t_0 < t_1 < \dots < t_N = 1$, by Theorem 1.1,

$$(2.3) \quad \begin{aligned} & \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in B(f, \rho) \right\} \\ & \leq \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(t_i)}{b(n)} \in \overline{B_\rho(f(t_i))}, i = 1, \dots, N \right\} \\ & \leq -\inf \left\{ \tilde{I}^\pi(z); z = (z_1, \dots, z_N), z_i \in \overline{B_\rho(f(t_i))}, i = 1, \dots, N \right\} \end{aligned}$$

where $B_\rho(x) = \{y \in \mathbf{E}; \|x - y\| < \rho\}$ for $x \in \mathbf{E}$, and $\tilde{I}^\pi(\cdot)$ is as in (1.8). We can choose partition π such that

$$(2.4) \quad \tilde{I}^\pi((f(t_1), \dots, f(t_N))) \geq \tilde{\Lambda}(f) - \epsilon/2.$$

By the lower semicontinuity of \tilde{I}^π , we can choose $\rho = \rho(f, \epsilon) > 0$ small enough such that

$$(2.5) \quad \begin{aligned} & \inf \left\{ \tilde{I}^\pi(z); z = (z_1, \dots, z_N), z_i \in \overline{B_\rho(f(t_i))}, 1 \leq i \leq N \right\} \\ & \geq \inf \left\{ \tilde{I}^\pi(z); z = (z_1, \dots, z_N), z_i \in B_{2\rho}(f(t_i)), 1 \leq i \leq N \right\} \\ & \geq \tilde{I}^\pi((f(t_1), \dots, f(t_N))) - \epsilon/2. \end{aligned}$$

So (2.2) follows from (2.3)-(2.5). If $\tilde{\Lambda}(f) = +\infty$, then we can similarly prove that (2.2) is still true.

Keeping (2.2) in mind, by a well-known standard argument we know that the upper bound holds for compact sets. Therefore, in order to complete the upper bound, it suffices to prove that $\left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \right\}$ is exponentially tight, that is, for every $L > 0$, there exists a compact set $K_L \subset C([0, 1], \mathbf{E})$, such that

$$(2.6) \quad \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in K_L^c \right\} \leq -L.$$

We will adapt Dembo and Zajic [16]'s argument to prove (2.6). However, a much more delicate estimate is needed when we prove (2.7) below.

By Theorem A in the Appendix A it suffices to prove

(i) For each rational $t \in [0, 1]$, $\left\{ \frac{\tilde{S}_n(t)}{b(n)} \right\}$ is exponentially tight, that is, for each $\alpha > 0$, there exists a compact set $K_\alpha \subset \mathbf{E}$ such that

$$\limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(t)}{b(n)} \in K_\alpha^c \right\} \leq -\alpha.$$

(ii) For each $\rho > 0$,

$$(2.7) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{1}{b(n)} \sup_{|t-s| < \delta} \|\tilde{S}_n(t) - \tilde{S}_n(s)\| \geq \rho \right\} = -\infty.$$

To this end. (i) follows from Lemma 2.4 and Lynch and Sethuraman [28, Lemma 2.6].

To prove (2.7). Noting Lemma 2.2, it is enough to prove

$$(2.8) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{1}{b(n)} \sup_{|t-s| < \delta} \|S_n(t) - S_n(s)\| \geq \rho \right\} = -\infty$$

for each $\rho > 0$.

In fact, fix $\rho > 0$, for $n \geq 1$ and $\delta > 0$,

$$(2.9) \quad \sup_{|t-s| < \delta} \|S_n(t) - S_n(s)\| \leq \max_{0 \leq k \leq n, 1 \leq m \leq [n\delta] + 1} \left\| \sum_{j=k+1}^{k+m} X_j \right\|.$$

For $0 < \delta \leq 1$, let $l = l(\delta) \geq 2$ be the unique integer satisfying $1/l < \delta \leq 1/(l-1)$. Then $n < ([n\delta] + 1)l$ for sufficiently large n . For such a large n , suppose

$$\|S_{k+m} - S_k\| = \left\| \sum_{j=k+1}^{k+m} X_j \right\| \geq b(n)\rho$$

for some $k, 0 \leq k \leq n$, and some $m, 1 \leq m \leq [n\delta] + 1$. Then there exists a unique integer $p, 0 \leq p \leq l-1$, such that

$$([n\delta] + 1)p \leq k < ([n\delta] + 1)(p+1).$$

Hence, there are two possibilities for $k+m$. One possibility is that

$$([n\delta] + 1)p \leq k+m < ([n\delta] + 1)(p+1),$$

in which case either $\|S_{k+m} - S_{([n\delta] + 1)p}\| \geq \frac{1}{3}b(n)\rho$, or $\|S_k - S_{([n\delta] + 1)p}\| \geq \frac{1}{3}b(n)\rho$. The second possibility is that

$$([n\delta] + 1)(p+1) \leq k+m < ([n\delta] + 1)(p+2),$$

in which case either $\|S_{k+m} - S_{([n\delta] + 1)(p+1)}\| \geq \frac{1}{3}b(n)\rho$, $\|S_{([n\delta] + 1)(p+1)} - S_{([n\delta] + 1)p}\| \geq \frac{1}{3}b(n)\rho$, or $\|S_k - S_{([n\delta] + 1)p}\| \geq \frac{1}{3}b(n)\rho$. In conclusion, we see that

$$(2.10) \quad \left\{ \max_{0 \leq k \leq n, 1 \leq m \leq [n\delta] + 1} \left\| \sum_{j=k+1}^{k+m} X_j \right\| \geq b(n)\rho \right\} \\ \subset \sum_{p=0}^l \left\{ \max_{1 \leq m \leq [n\delta] + 1} \left\| \sum_{j=p([n\delta] + 1) + 1}^{p([n\delta] + 1) + m} X_j \right\| \geq \frac{\rho}{3}b(n) \right\}.$$

Noting

$$P \left\{ \max_{1 \leq m \leq [n\delta] + 1} \left\| \sum_{j=p([n\delta] + 1) + 1}^{p([n\delta] + 1) + m} X_j \right\| \geq \frac{\rho}{3}b(n) \right\} \\ = P \left\{ \max_{1 \leq m \leq [n\delta] + 1} \left\| \sum_{j=1}^m X_j \right\| \geq \frac{\rho}{3}b(n) \right\},$$

by (2.9), (2.10) and Lemma 2.1,

$$\begin{aligned}
 (2.11) \quad & P \left\{ \frac{1}{b(n)} \sup_{|t-s|<\delta} \|S_n(t) - S_n(s)\| \geq \rho \right\} \\
 & \leq \sum_{p=0}^l P \left\{ \frac{1}{b(n)} \max_{1 \leq m \leq [n\delta]+1} \left\| \sum_{j=1}^m X_j \right\| \geq \frac{\rho}{3} \right\} \\
 & \leq 2(l+1)P \left\{ \frac{1}{b(n)} \left\| \sum_{j=1}^{[n\delta]+1} X_j \right\| \geq \frac{\rho}{6} \right\} \\
 & \leq 2(l+1)P \left\{ \frac{\|S_{[n\delta]}\|}{b(n)} \geq \frac{\rho}{12} \right\} + 2(l+1)P \left\{ \frac{\|X_1\|}{b(n)} \geq \frac{\rho}{12} \right\}.
 \end{aligned}$$

By (2.11), (1.4), Lemma 2.4 and Chebyshev's inequality, for all $0 < \delta \leq 1$,

$$\begin{aligned}
 (2.12) \quad & \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{1}{b(n)} \sup_{|t-s|<\delta} \|S_n(t) - S_n(s)\| \geq \rho \right\} \\
 & \leq \max \left\{ \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\|S_{[n\delta]}\|}{b(n)} \geq \frac{\rho}{12} \right\}, \right. \\
 & \quad \left. \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \|X_1\| \geq \frac{\rho}{12} b(n) \right\} \right\} \\
 & \leq -\frac{1}{\delta} \inf \left\{ \frac{1}{2} \|x\|_H^2; x \in H, \|x\| \geq \frac{\rho}{12} \right\}.
 \end{aligned}$$

Note that for $x \in H$,

$$(2.13) \quad \|x\| \leq \sigma \|x\|_H$$

where

$$\sigma = \sup_{\|g\| \leq 1, g \in \mathbf{E}^*} \left(\int_{\Omega} g^2(X_1) dP \right)^{1/2} < \infty.$$

(See Goodman, Kuebls and Zinn [22].) Taking limit $\delta \rightarrow 0$ in (2.12) implies (2.8). The upper bound is established.

Lower bound. To prove the lower bound, it suffices to prove that for each piecewise linear function $f \in C([0, 1], \mathbf{E})$, and each $\rho > 0$,

$$(2.14) \quad \liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in B(f, \rho) \right\} \geq -\tilde{\Lambda}(f).$$

(See also Schuette [31], Hu [23], etc.)

This can be easily reduced to proving for the case of a linear function $f : t \rightarrow tx, t \in [0, 1], x \in \mathbf{E}$ (see also Borovkov and Mogulskii [12]). So, we now focus on the case $f(t) = tx, t \in [0, 1]$, where $x \in \mathbf{E}$ is arbitrarily fixed.

For $\theta > 0, x \in \mathbf{E}$, let $N_\theta(x) = \{g \in C([0, 1], \mathbf{E}); \|g(1) - x\| < \theta\}$. Note that for $\rho > 0, \theta > 0$,

$$\overline{N_\theta(x)} \cap B(f, \rho) = \overline{N_\theta(x)} \setminus \{\overline{N_\theta(x)} \setminus B(f, \rho)\} \subset B(f, \rho).$$

Since the MDP for the partial sums $\left\{\frac{\tilde{S}_n(1)}{b(n)}\right\}$ yields

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in \overline{N_\theta(x)} \right\} \\ & \geq \liminf_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \left\| \frac{\tilde{S}_n(1)}{b(n)} - x \right\| < \theta \right\} \\ & \geq -\Lambda(x) = -\tilde{\Lambda}(f), \end{aligned}$$

in order to prove (2.14), it suffices to prove

$$(2.15) \quad \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log P \left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \in \overline{N_\theta(x)} \setminus B(f, \rho) \right\} < -\Lambda(x)$$

for a certain $\theta > 0$.

A little thought reveals that all sufficiently small $\theta > 0$ will do. Since $\overline{N_\theta(x)} \setminus B(f, \rho)$ is a closed set, we apply the upper bound result and estimate

$$\begin{aligned} & -(\text{L. H. S. of (2.15)}) \geq \inf \{ \tilde{\Lambda}(g); \|g - f\| \geq \rho, \|g(1) - x\| \leq \theta \} \\ & \geq \inf \{ \tilde{\Lambda}(g); \sup_{0 \leq t \leq 1} \|g(t) - (f(t) + tz)\| \geq \rho - \theta, z = g(1) - x, \|z\| \leq \theta \} \\ & = \inf_{\|z\| \leq \theta} \inf \{ \tilde{\Lambda}(g); \|g(t) - t(x + z)\| \geq \rho - \theta \text{ for some } 0 \leq t \leq 1, g(1) = x + z \} \\ & \geq \inf_{\|z\| \leq \theta} \inf_{0 \leq t \leq 1} \inf \{ \tilde{\Lambda}(g); y = g(t) - t(x + z), \|y\| \geq \rho - \theta, g(1) = x + z \} \\ & \geq \inf_{\|z\| \leq \theta} \inf_{0 \leq t \leq 1} \inf_{\|y\| \geq \rho - \theta} \left\{ t\Lambda \left((x + z) + \frac{y}{t} \right) + (1 - t)\Lambda \left((x + z) - \frac{y}{1 - t} \right) \right\} \\ & \geq \inf_{\|z\| \leq \theta} \inf_{\|y\|_H \geq \frac{\rho - \theta}{\sigma}} \inf_{0 \leq t \leq 1} \frac{1}{2} \left\{ t \left\| x + z + \frac{y}{t} \right\|_H^2 + (1 - t) \left\| x + z - \frac{y}{1 - t} \right\|_H^2 \right\} \\ & \geq \inf_{\|z\| \leq \theta} \inf_{\|y\|_H \geq \frac{\rho - \theta}{\sigma}} \inf_{0 \leq t \leq 1} \frac{1}{2} \left\{ \|x + z\|_H^2 + \left(\frac{1}{t} + \frac{1}{1 - t} \right) \|y\|_H^2 \right\} \\ & \geq \inf_{\|z\| \leq \theta} \left\{ \frac{1}{2} \|x + z\|_H^2 + 2 \left(\frac{\rho - \theta}{\sigma} \right)^2 \right\} \\ & = \inf_{\|z\| \leq \theta} \left\{ \Lambda(x + z) + 2 \left(\frac{\rho - \theta}{\sigma} \right)^2 \right\}, \end{aligned}$$

whose limit as $\theta \rightarrow 0$ is greater than or equal to $\Lambda(x) + 2 \left(\frac{\rho}{\sigma} \right)^2$ by the lower semi-continuity of the rate function Λ , where σ is as in (2.13).

This concludes that (2.15) indeed holds for sufficiently small $\theta > 0$. The proof of the lower bound is completed.

So far, we have proved that $\left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \right\}$ in $C([0, 1], \mathbf{E})$ satisfies an MDP with speed $\left\{ \frac{n}{b^2(n)} \right\}$ and the rate function $\tilde{\Lambda}$ defined by (1.2). It is easy to check that $(C([0, 1], \mathbf{E}), \|\cdot\|_\infty)$ is a closed subset of $(D([0, 1], \mathbf{E}), d_\infty)$, and therefore $\left\{ \frac{\tilde{S}_n(\cdot)}{b(n)} \right\}$ in $(D([0, 1], \mathbf{E}), d_\infty)$ also satisfies the MDP with speed $\left\{ \frac{n}{b^2(n)} \right\}$ and the rate function $\tilde{\Lambda}$ of (1.2) (see Dembo and Zeitouni [18, Lemma 4.1.5 (a)]. The exponential equivalence of

$\left\{\frac{\tilde{S}_n(\cdot)}{b(n)}\right\}$ and $\left\{\frac{S_n(\cdot)}{b(n)}\right\}$ in $(D([0, 1], \mathbf{E}), d_\infty)$, established in Lemma 2.2, implies now that $\left\{\frac{S_n(\cdot)}{b(n)}\right\}$ in $D([0, 1], \mathbf{E})$ satisfies the MDP with speed $\left\{\frac{n}{b^2(n)}\right\}$ and the rate function $\tilde{\Lambda}$ of (1.2) (see Dembo and Zeitouni [18, Theorem 4.2.13]). Theorem 1.2 is proved.

3. FUNCTIONAL LILS

Throughout this section, let $\beta(n) = \sqrt{2n \log \log n}$, $n \geq 3$. Based on Theorem 1.2 and Remark 1.2, we can conclude the Functional Laws of the Iterated Logarithm for the piecewise constant and piecewise linear functions, $\{S_n(\cdot); n \geq 1\}$ and $\{\tilde{S}_n(\cdot); n \geq 1\}$, respectively.

Theorem 3.1. *Suppose that (1.4) and the following condition*

$$\frac{S_n}{\beta(n)} \xrightarrow{P} 0$$

hold. Then with probability 1, the following sequence

$$\left\{\xi_n(\cdot) = \frac{S_n(\cdot)}{\beta(n)}\right\}_{n \geq 1}$$

is relatively compact in $D([0, 1], \mathbf{E})$, and the set of its limit points, $L(\omega)$, is precisely the compact set

$$\mathcal{K} = \left\{f \in D([0, 1], \mathbf{E}); 2\tilde{\Lambda}(f) \leq 1\right\}.$$

The same result holds for $\left\{\frac{\tilde{S}_n(\cdot)}{\beta(n)}\right\}_{n \geq 1}$.

Proof. It can be proved by the standard arguments, see, for example, the proof of Deuschel and Stroock [20, Theorem 1.4.1] or Dembo and Zajic [17, Corollary 1].

4. CONCLUDING REMARKS

We have viewed the trajectory problem as consisting of two major issues. First, what is the MDP for partial sums? For this we found good results in the literature, see, for instance, Chen [13, 15], Ledoux [26]. Second, how to pass a result from the partial sum to the whole trajectory, now that the result holds for partial sums? The latter issue is treated carefully in Section 2.

We have traced all the proofs and seen that, once the MDP for partial sums is assumed, the original assumption (1.4) is rarely quoted in settling the second issue. This, among other things, suggests that any partial sum result may well remain true for the corresponding trajectory process. Let us illustrate use of such an idea by extending a partial sum result to the trajectory setting, the results in Proposition 4.1 and Proposition 4.2 below. In the proof, we will list all occasions of quoting the original assumption, (1.4). It should be pointed out that (1.5) is a necessary condition for (1.6) to hold.

Proposition 4.1. *Let $b(n) = n^p$ ($1/2 < p < 1$). Suppose that (1.5) and*

$$E \exp(\beta \|X_1\|^\alpha) < \infty \text{ for some } 2 - 1/p < \alpha < 1 \text{ and some } \beta = \beta(\alpha) > 0$$

hold. Then Theorem 1.1 and Theorem 1.2 remain true.

Proof. By Chen [13, Theorem 2]; [15, Theorem 1] or Jiang [24], $\left\{\frac{S_n}{b(n)}\right\}$ satisfies an MDP with speed $\left\{\frac{n}{b^2(n)}\right\}$ and rate function Λ as in (1.1).

Again by tracing the proofs of Chen [13, Theorem 2]; [15, Theorem 1], we can show that Lemma 2.3 is still true under the new assumption instead of (1.4).

Excluding Lemma 2.3, use was made of (1.4) only in the proofs of Lemma 2.2 and (2.12) via the following estimate

$$P\{\|X_1\| > b(n)\delta\} \leq \exp(-\beta\delta b(n))E\exp(\beta\|X_1\|)$$

for $\delta > 0$, where β is as in (1.4). The inequality above ensures the desired estimate

$$(4.1) \qquad \limsup_{n \rightarrow \infty} \frac{n}{b^2(n)} \log(nP\{\|X_1\| > b(n)\delta\}) = -\infty$$

for each $\delta > 0$. Therefore, all that remains is to show that (4.1) is valid under the new assumption, weaker than (1.4). Indeed, for each $\delta > 0$, by Chebyshev's inequality,

$$nP\{\|X_1\| > \delta b(n)\} \leq n \exp(-\beta\delta^\alpha b^\alpha(n)) E \exp(\beta\|X_1\|^\alpha),$$

which implies (4.1). Proof is completed.

Proposition 4.2. *Let $b(n) = \sqrt{2n \log \log n}$, $n \geq 3$. Suppose that (1.5) and*

$$E \exp(\beta\|X_1\|^\alpha) < \infty \text{ for some } 0 < \alpha < 1 \text{ and some } \beta = \beta(\alpha) > 0$$

hold. Then Theorem 1.1, Theorem 1.2 and Theorem 3.1 hold.

Proof. Its proof is similar to that of Proposition 4.1.

Remark 4.1. Proposition 4.2 has improved Theorem 3.1 by weakening the exponential integrability assumption, (1.4).

APPENDIX A

Let (\mathcal{X}, d) be a Polish space and \mathcal{Y} denote the Polish space of continuous functions from $[0, 1]$ to \mathcal{X} equipped with the metric $d_\infty(f, g) = \sup_{0 \leq t \leq 1} d(f(t), g(t))$. Let $\{a(n)\}$ be a positive sequence satisfying $a(n) \rightarrow 0$ as $n \rightarrow \infty$.

A sequence of probability measures $\{\mu_n; n \geq 1\}$ on \mathcal{Y} is said to be exponentially tight with speed $\{a(n)\}$ if for every $L > 0$, there exists a compact set $K_L \subset \mathcal{Y}$ such that

$$(A.1) \qquad \limsup_{n \rightarrow \infty} a(n) \log \mu_n \{K_L^c\} \leq -L,$$

where K_L^c means the complement of K_L .

Theorem A. *A sequence of probability measures $\{\mu_n; n \geq 1\}$ on \mathcal{Y} is exponentially tight with speed $\{a(n)\}$ if;*

(i) *For each rational $t \in [0, 1]$, the sequence $\{\mu_n(t); n \geq 1\}$ of laws induced by the projection $f(\cdot) \rightarrow f(t) : \mathcal{Y} \rightarrow \mathcal{X}$ is exponentially tight in (\mathcal{X}, d) , that is, for each $\alpha > 0$, there exists a compact set $L_\alpha \subset \mathcal{X}$ such that*

$$\limsup_{n \rightarrow \infty} a(n) \log \mu_n \{f(t) \in L_\alpha^c\} \leq -\alpha.$$

(ii) *For all $\rho > 0$,*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} a(n) \log \mu_n \{\{f; \omega_f(\delta) \geq \rho\}\} = -\infty,$$

where for each $f \in \mathcal{Y}$ and all $\delta > 0$,

$$\omega_f(\delta) = \sup_{|t-s|<\delta} d(f(t), f(s))$$

denotes the modulus of continuity of f .

Proof. This is essentially the Lemma A.2 of Dembo and Zajic [16].

APPENDIX B

Let $\tilde{\Lambda}$ be defined by (1.2). The purpose of this section is to show that $\tilde{\Lambda}$, defined on $(D([0, 1], \mathbf{E}), d_\infty(\cdot, \cdot))$, has compact level sets. Since $(C([0, 1], \mathbf{E}), \|\cdot\|_\infty)$ is a closed subspace of $(D([0, 1], \mathbf{E}), d_\infty(\cdot, \cdot))$, it suffices for us to show that $\tilde{\Lambda}$ restricted to $(C([0, 1], \mathbf{E}), \|\cdot\|_\infty)$ has compact level sets. Throughout this section, we denote by λ the Lebesgue measure on $[0, 1]$. Note that Λ has compact level sets under the integrability condition (1.4).

Lemma B.1. *Let $K_a = \{\varphi \in C([0, 1]; \mathbf{E}); \tilde{\Lambda}(\varphi) \leq a\}$ for $a > 0$, and let $\dot{K}_a = \{g \in L^1([0, 1], \mathbf{E}); \varphi(t) = \int_0^t g(s)ds \text{ for } t \in [0, 1], \varphi \in K_a\}$. Then \dot{K}_a is $\|\cdot\|_H$ -uniformly integrable, that is to say*

$$(B.1) \quad \lim_{k \rightarrow \infty} \sup_{g \in \dot{K}_a} \int_{\{\|g\|_H \geq k\}} \|g\|_H d\lambda = 0.$$

Proof. Given $g \in \dot{K}_a$, for each $k > 0$, it follows from the Cauchy-Schwartz inequality that

$$(B.2) \quad \int_{\{\|g\|_H \geq k\}} \|g\|_H d\lambda \leq \frac{2}{k} \tilde{\Lambda}(\varphi) \leq \frac{2}{k} a,$$

where $\varphi(t) = \int_0^t g(s)ds$ for $t \in [0, 1]$. (B.1) follows from (B.2). The lemma is proved.

Lemma B.2 (Lower semicontinuity of $\tilde{\Lambda}$). *If $\|\varphi_n - \varphi\|_\infty \rightarrow 0$ as $n \rightarrow \infty$, then $\liminf_{n \rightarrow \infty} \tilde{\Lambda}(\varphi_n) \geq \tilde{\Lambda}(\varphi)$.*

Proof. It suffices to consider $\liminf_{n \rightarrow \infty} \tilde{\Lambda}(\varphi_n) = b < \infty$. By passing to a subsequence, we may and will assume that $\lim_{n \rightarrow \infty} \tilde{\Lambda}(\varphi_n) = b$, and $\tilde{\Lambda}(\varphi_n) \leq b + 1$ for all n .

Let $g_n \in L^1([0, 1], \mathbf{E})$ such that $g_n(t) \in H$ and $\varphi_n(t) = \int_0^t g_n(s)ds$ for $t \in [0, 1]$.

We shall show that φ is $\|\cdot\|_H$ -absolutely continuous: that is, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that $n \in \mathbf{N}$, $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n \leq 1$, $\sum(t_i - s_i) < \delta$ imply

$$(B.3) \quad \sum_{i=1}^n \|\varphi(t_i) - \varphi(s_i)\|_H < \varepsilon.$$

To prove (B.3). Given $\varepsilon > 0$, by Lemma B.1, there exists $\delta = \delta(\varepsilon) > 0$ such that for all n , $\int_A \|g_n\|_H d\lambda < \varepsilon$ whenever $\lambda(A) < \delta$. In particular, if $s_i < t_i$ and $\sum |t_i - s_i| < \delta$, then

$$(B.4) \quad \sum \|\varphi_n(t_i) - \varphi_n(s_i)\|_H < \varepsilon.$$

Taking account of (B.4), the lower semicontinuity of Λ and $\|\varphi_n - \varphi\|_\infty \rightarrow 0$, we can obtain

$$(B.5) \quad \varepsilon \geq \liminf_{n \rightarrow \infty} \sum (2\Lambda(\varphi_n(t_i) - \varphi_n(s_i)))^{1/2} \geq \sum (2\Lambda(\varphi(t_i) - \varphi(s_i)))^{1/2}.$$

Note that $\varphi(0) = \lim_{n \rightarrow \infty} \varphi_n(0) = 0 \in H$. (B.5), together with the definition of Λ , implies that $\varphi(s_i), \varphi(t_i) \in H$. In return, we have $\varphi(t) \in H$ for all $t \in [0, 1]$ and

$$(B.6) \quad \sum \|\varphi(t_i) - \varphi(s_i)\|_H \leq \varepsilon$$

which means that φ is $\|\cdot\|_H$ -absolutely continuous. Note that (B.6) yields also that $\int_0^1 \|\varphi(s)\|_H ds < \infty$.

We next show that there exists $g \in L^1([0, 1], \mathbf{E})$ such that $g(t) \in H$ and $\varphi(t) = \int_0^t g(s)ds$ for $t \in [0, 1]$. To this end, define on $([0, 1], \mathcal{B}, \lambda)$, where \mathcal{B} is the Borel σ -algebra of $[0, 1]$, the H -valued martingale (h_n, \mathcal{F}_n) , where

$$h_n = \sum_{j=1}^{2^n} 2^n \left[\varphi\left(\frac{j}{2^n}\right) - \varphi\left(\frac{j-1}{2^n}\right) \right] \mathbf{1}_{[(j-1)/2^n, j/2^n)}$$

and $\mathcal{F}_n = \sigma\left(\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right]; 1 \leq j \leq 2^n\right)$.

Because φ is $\|\cdot\|_H$ -absolutely continuous, it is of $\|\cdot\|_H$ -bounded variation; that is, there exists a positive constant $M < \infty$, such that if $n \in \mathbf{N}$ and $0 \leq t_0 < t_1 < \dots < t_n \leq 1$, then

$$(B.7) \quad \sum_{i=1}^n \|\varphi(t_i) - \varphi(t_{i-1})\|_H \leq M.$$

Since for $n \in \mathbf{N}$,

$$E\|h_n\|_H = \sum_{j=1}^{2^n} \left\| \varphi\left(\frac{j}{2^n}\right) - \varphi\left(\frac{j-1}{2^n}\right) \right\|_H,$$

it follows from (B.7) that $\sup_n E\|h_n\|_H < \infty$. By the well-known martingale convergence theorem, there exists $g \in L^1([0, 1], \mathbf{E})$ satisfying $g(t) \in H$ for $t \in [0, 1]$ and

$\int_0^1 \|g(s)\|_H ds < \infty$ such that $\lim_{n \rightarrow \infty} \|h_n - g\|_H = 0$ a.e. Next, we shall show that

$\varphi(t) = \int_0^t g(s)ds$ for $t \in [0, 1]$. First of all, we shall show that

$$(B.8) \quad \lim_{n \rightarrow \infty} E\|h_n - g\|_H = 0.$$

To prove (B.8), it is enough to show $\{h_n\}$ is $\|\cdot\|_H$ -uniformly integrable, that is

$$(B.9) \quad \lim_{\rho \rightarrow \infty} \sup_n \int_{\{\|h_n\|_H \geq \rho\}} \|h_n\|_H d\lambda = 0.$$

To show this, let $\Delta_j = \varphi\left(\frac{j}{2^n}\right) - \varphi\left(\frac{j-1}{2^n}\right)$, $j = 1, \dots, 2^n$, then

$$C \triangleq \sup_n \sum_{j=1}^{2^n} \|\Delta_j\|_H < \infty$$

and

$$(B.10) \quad 2^{-n} \text{card}\{j : 2^n \|\Delta_j\|_H \geq \rho\} \leq \rho^{-1} \sum_{j=1}^{2^n} \|\Delta_j\|_H \leq \frac{C}{\rho}.$$

However

$$(B.11) \quad \int_{\{\|h_n\|_H \geq \rho\}} \|h_n\|_H d\lambda = \sum_{\{j: 2^n \|\Delta_j\|_H \geq \rho\}} \|\Delta_j\|_H.$$

Now, (B.9) follows from (B.10), (B.11) and the $\|\cdot\|_H$ -absolute continuity of φ , (B.3). Consequently, by the definition of Bochner integral, for $0 \leq j < k \leq 2^n$,

$$(B.12) \quad \int_{j/2^n}^{k/2^n} g(s) ds = \lim_{l \rightarrow \infty} \int_{j/2^n}^{k/2^n} h_l(s) ds = \varphi\left(\frac{k}{2^n}\right) - \varphi\left(\frac{j}{2^n}\right).$$

Recall that the Bochner integral $h(u) = \int_0^u g(s) ds : [0, 1] \rightarrow \mathbf{E}$ is continuous (see Diestel and Uhl [21, Theorem II.2.4]), and therefore, by (B.12) and the continuity of φ , for $0 \leq s < t \leq 1$, we have

$$\int_s^t g(s) ds = \varphi(t) - \varphi(s).$$

In particular, for $t \in [0, 1]$, we have

$$\varphi(t) = \int_0^t g(s) ds.$$

Finally, we shall show that $b \geq \tilde{\Lambda}(\varphi)$. Let $\Pi : 0 = t_0 < t_1 < \dots < t_N = 1$ be a partition of $[0, 1]$, where $\|\Pi\| \triangleq \max_{1 \leq i \leq N} |t_i - t_{i-1}|$ will be taken to be sufficiently small. Note that $\tilde{\Lambda}(\varphi_n) = \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \Lambda(g_n) d\lambda$. By Jensen's inequality,

$$\int_{t_{i-1}}^{t_i} \Lambda(g_n) d\lambda \geq (t_i - t_{i-1}) \Lambda\left(\frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} g_n d\lambda\right).$$

Therefore

$$\tilde{\Lambda}(\varphi_n) \geq \sum_{i=1}^N (t_i - t_{i-1}) \Lambda\left(\frac{\varphi_n(t_i) - \varphi_n(t_{i-1})}{t_i - t_{i-1}}\right).$$

Using the lower semicontinuity of Λ , we obtain

$$(B.13) \quad b = \liminf_{n \rightarrow \infty} \tilde{\Lambda}(\varphi_n) \geq \sum_{i=1}^N (t_i - t_{i-1}) \Lambda\left(\frac{\varphi(t_i) - \varphi(t_{i-1})}{t_i - t_{i-1}}\right) = \tilde{\Lambda}(\varphi_\Pi)$$

for any partition Π , where φ_Π is defined as follows

$$\varphi_\Pi(t) = \int_0^t g_\Pi(s) ds, \quad t \in [0, 1],$$

where

$$g_\Pi(t) = \frac{\varphi(t_i) - \varphi(t_{i-1})}{t_i - t_{i-1}} \quad \text{for } t \in [t_{i-1}, t_i)$$

and $g_\Pi(1) = \varphi(1)$.

Let Π_m be a partition such that $\|\Pi_m\| \leq \frac{1}{m}$. Define

$$g_{\Pi_m}(t) = \frac{\varphi\left(t_i^{(m)}\right) - \varphi\left(t_{i-1}^{(m)}\right)}{t_i^{(m)} - t_{i-1}^{(m)}} \quad \text{for } t \in \left[t_{i-1}^{(m)}, t_i^{(m)}\right)$$

and $g_{\Pi_m}(1) = \varphi(1)$, where $\Pi_m : 0 = t_0^{(m)} < t_1^{(m)} < \dots < t_{m_N}^{(m)} = 1$. Since $\int_0^1 \|g(s)\|_H ds < \infty$, $\lim_{m \rightarrow \infty} \|g_{\Pi_m} - g\|_H = 0$ a.e. Furthermore, $\lim_{m \rightarrow \infty} \|g_{\Pi_m} - g\| = 0$ a.e. Consequently, taking account of the lower semicontinuity of Λ , we have $\liminf_{m \rightarrow \infty} \Lambda(g_{\Pi_m}) \geq \Lambda(g)$ a.e., which, as well as Fatou's lemma, implies

$$(B.14) \quad \liminf_{m \rightarrow \infty} \int_0^1 \Lambda(g_{\Pi_m}) d\lambda \geq \int_0^1 \liminf_{m \rightarrow \infty} \Lambda(g_{\Pi_m}) d\lambda \geq \int_0^1 \Lambda(g) d\lambda = \tilde{\Lambda}(\varphi).$$

From (B.13) and (B.14) it follows that

$$b \geq \liminf_{m \rightarrow \infty} \tilde{\Lambda}(g_{\Pi_m}) \geq \tilde{\Lambda}(\varphi),$$

which proves the desired results. The lemma is proved.

Lemma B.3. For any $a > 0$, $K_a = \{\varphi : \tilde{\Lambda}(\varphi) \leq a\}$ is compact in $(C[0, 1], \|\cdot\|_\infty)$.

Proof. Note first that: if $A \subset C([0, 1]; \mathbf{E})$ is such that

- (i) There exist a compact set $K \subset \mathbf{E}$ such that $\varphi(t) \in K$ for all $t \in [0, 1], \varphi \in A$;
- (ii) $\limsup_{\delta \rightarrow 0} \omega_\varphi(\delta) = 0$, where $\omega_\varphi(\delta) = \sup_{|t-s| < \delta} \|\varphi(t) - \varphi(s)\|$, then A is compact

in $(C([0, 1]; \mathbf{E}), \|\cdot\|_\infty)$ (see also de Acosta [3, p. 88]).

Given $\varphi \in K_a$, let $g \in L^1([0, 1], \mathbf{E})$ such that $\varphi(t) = \int_0^t g(s) ds$ for $t \in [0, 1]$, then

$$\|\varphi(t) - \varphi(s)\| = \left\| \int_s^t g(\tau) d\tau \right\| \leq \int_s^t \|g(\tau)\| d\tau,$$

which, together with Lemma B.1, yields (ii) for $A = K_a$.

Given $\varphi \in K_a$ again, let $g \in L^1([0, 1], \mathbf{E})$ such that $\varphi(t) = \int_0^t g(s) ds$ for $t \in [0, 1]$.

For any $t \in (0, 1]$, by Jensen's inequality

$$\Lambda(\varphi(t)) = t^2 \Lambda\left(\frac{1}{t} \int_0^t g(s) ds\right) \leq t \int_0^t \Lambda(g(s)) ds \leq \tilde{\Lambda}(\varphi) \leq a$$

This proves (i) for $A = K_a$, with $K = \{x \in \mathbf{E}; \Lambda(x) \leq a\}$, where $\{x \in \mathbf{E}; \Lambda(x) \leq a\}$ is compact in $(\mathbf{E}, \|\cdot\|)$. Now the compactness of K_a follows from the above arguments and the lower semicontinuity of $\tilde{\Lambda}$ (Lemma B.2). The proof is completed.

From Lemma B.2 and Lemma B.3, we can obtain following theorem.

Theorem B.1. Let $\tilde{\Lambda}$ be defined by (1.2). Under condition (1.4), $\tilde{\Lambda}$ is a rate function on $(C([0, 1]; \mathbf{E}), \|\cdot\|_\infty)$.

ACKNOWLEDGMENTS

We would like to thank Professor Krzysztof Burdzy and Professor Amir Dembo for their comments. We are grateful to Professor Miguel A. Arcones for sending us his preprint. We are indebted to a referee who suggested Lemma 2.3 as well as the proof of Lemma 2.4, which led to the present improved version of the manuscript.

REFERENCES

- [1] de Acosta, A.: Upper bounds for large deviations of dependent random vectors, *Z. Wahrsch. Verw. Gebiete* 60, 551-565(1985). MR **87f**:60036
- [2] de Acosta, A.: Moderate deviations and associated Laplace approximations for sums of independent random vectors, *Trans. Amer. Math. Soc.* 329, 357-374(1992). MR **92e**:60053
- [3] de Acosta, A.: Large deviations for vector-valued Lévy processes, *Stochastic Process. Appl.* 51, 75-115 (1994). MR **96b**:60060
- [4] de Acosta, A.: Moderate deviations for empirical measures of Markov chains: lower bounds, *Ann. Probab.* 25, 259-284 (1997). MR **98f**:60049
- [5] de Acosta, A., Chen, Xia : Moderate deviations for empirical measures of Markov chains: upper bounds, *J. Theoret. Probab.* 11, 1075-1110 (1998) MR **99k**:60069
- [6] Araujo, A., Gine, E.: *The Central Limit Theorem for Real and Banach Valued Random Variables*, New York: Wiley, 1980. MR **83e**:60003
- [7] Arcones, M.A.: The large deviation principle for stochastic processes, Preprint, 2000.
- [8] Billingsley, P.: *Convergence of Probability Measures*, New York: Wiley 1968. MR **38**:1718
- [9] Bolthausen, E.: Laplace approximations for sums of independent random vectors, *Probab. Theory Related Fields* 72, 305-318(1986). MR **88b**:60075
- [10] Borovkov, A. A.: Boundary value problems for random walks and large deviations in function spaces, *Theory Probab. Appl.* 12, 575-595(1967). (Russian original, MR **39**:4906)
- [11] Borovkov, A. A., Mogulskii, A. A.: Probabilities of large deviations in topological space I. *Siberian Math. J.* 19, 697-709(1978) (Russian original, MR **80c**:60045)
- [12] Borovkov, A. A., Mogulskii, A. A.: Probabilities of large durations in topological space II. *Siberian Math. J.* 21, 653-664(1980). (Russian original, MR **82c**:60049)
- [13] Chen, Xia: On the lower bound of the moderate deviations of i.i.d. random variables in a Banach space, *Chinese Ann. Math.* 11A, 621-629(1990). (in Chinese).
- [14] Chen, Xia: Probabilities of moderate deviations for \mathbf{B} -valued independent random vectors, *Chinese J. Contemporary Math.* 11, 381-393(1990).
- [15] Chen, Xia: The moderate deviations of independent random vectors in a Banach space, *Chinese J. Appl. Probab. and Statist.* 7, 24-32(1991). MR **93m**:60020
- [16] Dembo, A., Zajic, T.: Large deviations: From empirical mean and measure to partial sums process, *Stochastic Process. Appl.* 57, 191-224(1995) MR **96m**:60064
- [17] Dembo, A., Zajic, T.: Uniform large and moderate deviations for functional empirical processes, *Stochastic Process. Appl.* 67, 195-211(1997) MR **98e**:60013
- [18] Dembo, A., Zeitouni, O.: *Large Deviations Techniques and Applications*, Boston: Jones and Bartlett, 1993. MR **95a**:60034
- [19] Deshayes, J., Picard, D.: Grandes et moyennes déviations pour le marches aleatoires, *Astérisque* 68, 53-71 (1979).
- [20] Deuschel, J. D., Stroock, D. W.: *Large Deviations*, Boston: Academic Press 1989. MR **90h**:60026
- [21] Diestel, J., Uhl, J.: *Vector Measures*, American Mathematical Society, Providence, 1977. MR **56**:12216
- [22] Goodman, V., Kuelbs, J., Zinn, J.: Some results on the LIL in Banach space with applications to weighted empirical processes, *Ann. Probab.* 9, 713-752 (1981). MR **82m**:60011
- [23] Hu, Yijun: Large deviations for trajectories of sums of dependent random variables, *Chinese J. of Contemporary Math*, 19, 149-158(1998). MR **99i**:60055
- [24] Jiang, Tiefeng: Moderate deviations for double arrays with applications to moving average processes, Gaussian processes and Cesaro's sums, Preprint (1999).
- [25] Karatzas, I., Shreve, S. E., *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag, 1991. MR **92h**:60127
- [26] Ledoux, M.: Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi, *Probab. and Statist. Ann. Inst. Henri Poincaré*, 28, 267-280(1992). MR **93k**:60017
- [27] Ledoux, M., Talagrand, M.: *Probability in Banach Spaces (Isoperimetry and Processes)*, Springer-Verlag, 1991. MR **93c**:60001
- [28] Lynch, J., Sethuraman, J.: Large deviations for processes with independent increments, *Ann. Probab.* 15, 610-627(1987). MR **88m**:60076
- [29] Mogulskii, A. A.: Large deviations in the space $C[0,1]$ for sums given on a finite Markov chain, *Siberian Math. J.* 15, 43-53(1974). (Russian original, MR **49**:9914)

- [30] Mogulskii, A. A.: Large deviations for trajectories of multi-dimensional random walks, Theory Probab. Appl. 21, 300-315(1976). (Russian original, MR 54:8810)
- [31] Schuette, P. H.: Large deviations for trajectories of sums of independent random variables, J. Theor. Probab. 7, 3-45(1994). MR 94m:60063
- [32] Varadhan, S. R. S.: Asymptotic probabilities and differential equations, Comm. Pure Appl. Math. 19, 261-286(1966) MR 34:3083
- [33] Wu, Liming: Some general methods of large deviations and applications, Preprint (1991), in <(Habilitation a diriger des recherches)>, Laboratoire de Probabilités de Paris VI, 1993.

DEPARTMENT OF MATHEMATICS, WUHAN UNIVERSITY, WUHAN, HUBEI 430072, PEOPLE'S REPUBLIC OF CHINA

E-mail address: yijunhu@public.wh.hb.cn

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MARYLAND 20742

E-mail address: tyl@math.umd.edu

WEIERSTRASS FUNCTIONS WITH RANDOM PHASES

YANICK HEURTEAUX

ABSTRACT. Consider the function

$$f_{\theta}(x) = \sum_{n=0}^{+\infty} b^{-n\alpha} g(b^n x + \theta_n),$$

where $b > 1$, $0 < \alpha < 1$, and g is a non-constant 1-periodic Lipschitz function. The phases θ_n are chosen independently with respect to the uniform probability measure on $[0, 1]$. We prove that with probability one, we can choose a sequence of scales $\delta_k \searrow 0$ such that for every interval I of length $|I| = \delta_k$, the oscillation of f_{θ} satisfies $\text{osc}(f_{\theta}, I) \geq C|I|^{\alpha}$. Moreover, the inequality $\text{osc}(f_{\theta}, I) \geq C|I|^{\alpha+\varepsilon}$ is almost surely true at every scale. When b is a transcendental number, these results can be improved: the minoration $\text{osc}(f_{\theta}, I) \geq C|I|^{\alpha}$ is true for every choice of the phases θ_n and at every scale.

1. INTRODUCTION

The function

$$w(x) = \sum_{n=0}^{+\infty} b^{-n\alpha} \cos(2\pi b^n x),$$

where $b > 1$ and $\alpha \leq 1$, is probably one of the most famous continuous nowhere differentiable functions. This function was introduced by Weierstrass. He proved that w is nowhere differentiable for some of these values b and α . A few years later, Hardy gave the proof for every $b > 1$ and every $\alpha \leq 1$ (see [5]).

More generally, one can consider the function

$$(1.1) \quad f(x) = \sum_{n=0}^{+\infty} b^{-n\alpha} g(b^n x),$$

where g is a 1-periodic Lipschitz function, $1 < b < +\infty$ and $0 < \alpha < 1$. Such a function will be called a Weierstrass function, and it is easy to prove that f is of class C^{α} (see, for example, [3]). Note that the regularity property of f is more subtle when $\alpha = 1$. In that case, f lies in the Zygmund class but is often not Lipschitz.

A famous conjecture states that the Hausdorff dimension of the graph of f is equal to $2 - \alpha$ (this is the biggest value that one can hope for). There are many papers that give general support to this conjecture (see, for example, [11], [10], [6], [7], [9], [14], [15]). Of course, one cannot expect this conjecture to be true in

Received by the editors July 8, 2002.

2000 *Mathematics Subject Classification*. Primary 26A27, 28A80, 37A05; Secondary 60F20.

Key words and phrases. Weierstrass functions, almost periodic functions, oscillations, fractal dimension.

general. Suppose for instance that g is of the form $g(x) = \varphi(x) - b^{-\alpha}\varphi(bx)$ with φ a smooth function. Then $f = \varphi$, and the graph has dimension 1. Thus, the genuine question is whether the conjecture is true when g is not of the above form. In recent work with Thierry Bousch, we proved the following result about the oscillations of the function f , which supports the conjecture.

Theorem 1.1 ([2], [1]). *Let g be a 1-periodic Lipschitz function, $1 < b < +\infty$ and $0 < \alpha < 1$. Define f using formula (1.1). There are only two possible mutually exclusive cases:*

$$(i) \ f \text{ is Lipschitz and } \|f'\|_{\infty} \leq \frac{\|g'\|_{\infty}}{b^{1-\alpha}-1},$$

or

$$(ii) \text{ there exists a constant } C > 0 \text{ such that for every interval } I \text{ of length } |I| \leq 1, \\ (1.2) \quad \operatorname{osc}(f, I) = \sup_I(f) - \inf_I(f) \geq C|I|^{\alpha}.$$

Moreover, the set of functions g such that (ii) is satisfied is a dense open subset of the space of 1-periodic Lipschitz functions (equipped with its standard norm).

The proof of this result is based on the following functional equation:

$$(1.3) \quad f(x) = g(x) + b^{-\alpha}f(bx),$$

which is satisfied by the Weierstrass function f . Let us also recall that conclusion (1.2) is often present in the literature about Weierstrass-type functions. In [14], sufficient conditions on g are given which ensure that (1.2) is satisfied. Assuming some stronger properties on g , Kaplan et al. ([9]) prove a result similar to Theorem 1.1 about the behavior of f . Finally, in [7], Hu and Lau prove a conclusion slightly different from (1.2) in the case where the Weierstrass-Mandelbrot function associated with f is not identically equal to zero.

It is well known that conclusion (1.2) ensures that the box-counting dimension of the graph of f is greater than $2 - \alpha$ (see, for example, [3]). On the other hand, McMullen shows in [12] that there are self-affine functions of class C^{α} satisfying (1.2) but whose graphs have Hausdorff dimension strictly less than $2 - \alpha$ (see also [14]). These functions are not Weierstrass-type functions and do not refute the conjecture.

The purpose of this paper is to study the local oscillation behavior of Weierstrass functions with phases. Such a function will be defined by the formula

$$(1.4) \quad f_{\theta}(x) = \sum_{n=0}^{+\infty} b^{-n\alpha} g(b^n x + \theta_n),$$

where g is a 1-periodic Lipschitz function, $1 < b < +\infty$, $0 < \alpha < 1$ and $\theta = (\theta_n)_{n \geq 0}$ is a sequence of phases satisfying $\theta_n \in [0, 1]$. It is easy to check that the function f_{θ} is still Hölder continuous on \mathbb{R} with exponent α (see, for example, [3] or [8]). More precisely, one can prove that the Hölder constant C of f_{θ} satisfies

$$C \leq \left[\frac{\|g'\|_{\infty}}{1 - b^{\alpha-1}} + \frac{\operatorname{osc}(g)}{1 - b^{-\alpha}} \right],$$

where $\operatorname{osc}(g) = \sup_{\mathbb{R}}(g) - \inf_{\mathbb{R}}(g)$ is the global oscillation of the function g . In particular, this constant does not depend on θ .

When b is an integer, the function f_{θ} is 1-periodic. In the other cases we can only ensure that f_{θ} is almost periodic (as a uniform limit of a sequence of almost

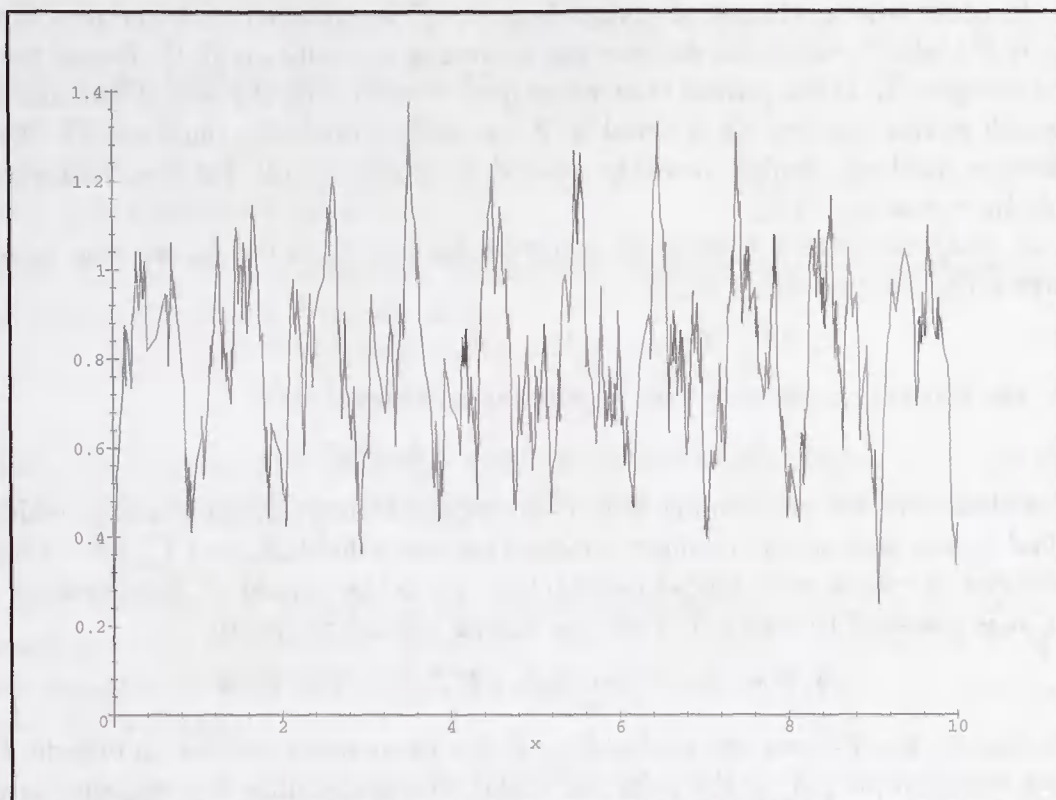


FIGURE 1. Graph of $f_\theta(x)$ with $g(x) = \text{dist}(x, \mathbb{Z})$, $b = 2.1$ and $\alpha = 0.5$

periodic functions). More precisely, we can establish the following property, which states that the almost periodicity property does not depend on the sequence of phases θ . Such a property will be useful in the proof of the main theorem (see Lemma 4.2).

Proposition 1.2 (Equi almost periodicity). *The family of functions $(f_\theta)_{\theta \in [0,1]^{\mathbb{N}}}$ satisfies the following property:*

$$\forall \varepsilon > 0, \exists \ell > 0; \forall \theta \in [0,1]^{\mathbb{N}}, \forall \delta \in \mathbb{R}, \exists a \in [\delta, \delta + \ell); \|\tau_a f_\theta - f_\theta\|_\infty \leq \varepsilon$$

($\tau_a f_\theta$ is defined by $\tau_a f_\theta(x) = f_\theta(x + a)$).

Proof. This proposition is a consequence of the compactness of the set $[0,1]^{\mathbb{N}}$. Let $\varepsilon > 0$ and $\theta \in [0,1]^{\mathbb{N}}$. We know that the function f_θ is almost periodic. So, we can find $\ell_\theta > 0$ such that for all $\delta \in \mathbb{R}$, there exists $a_\theta \in [\delta, \delta + \ell_\theta)$ with $\|\tau_{a_\theta} f_\theta - f_\theta\|_\infty \leq \varepsilon/3$. The application $\theta \in [0,1]^{\mathbb{N}} \mapsto f_\theta \in \mathcal{C}_b(\mathbb{R})$ being continuous (when the set $\mathcal{C}_b(\mathbb{R})$ of bounded continuous functions is endowed with the norm $\|\cdot\|_\infty$), we can find $\theta^1, \dots, \theta^n \in [0,1]^{\mathbb{N}}$ such that the set $\{f_\theta\}$ is covered by the balls $B(f_{\theta^k}, \varepsilon/3)$. Let $\ell = \max(\ell_{\theta^1}, \dots, \ell_{\theta^n})$. If $\delta \in \mathbb{R}$ and if $f_\theta \in B(f_{\theta^k}, \varepsilon/3)$, we obtain $\|\tau_{a_{\theta^k}} f_\theta - f_\theta\|_\infty \leq \varepsilon$. The conclusion follows if we note that $a_{\theta^k} \in [\delta, \delta + \ell)$. \square

As we can see in Figure 1, the function f_θ seems to be irregular, and we would like to know if a minoration similar to (1.2) is true for Weierstrass functions with phases. In the general case, such a minoration seems to be difficult to obtain for each value of the sequence $\theta = (\theta_n)_{n \geq 0}$. That is the reason why we propose to

consider the phases θ_n as independent random variables uniformly distributed in $[0, 1]$. In other words, the set of phases $\Omega = [0, 1]^{\mathbb{N}}$ is endowed with the probability measure $\mathbb{P} = dx^{\otimes \mathbb{N}}$, where dx denotes the Lebesgue measure on $[0, 1]$. Recall that in such a context, B. Hunt proved that when $g(x) = \cos x$, the Hausdorff dimension of the graph of the function f_θ is equal to $2 - \alpha$ with probability one (see [8]). Using a different method, Szulga recently proved a similar result for the Weierstrass-Mandelbrot process ([16]).

If we want to write a functional equation for the function f_θ , we also have to introduce the shift operator on Ω :

$$(1.5) \quad T : (\theta_n)_{n \geq 0} \in \Omega \mapsto (\theta_{n+1})_{n \geq 0} \in \Omega.$$

Then, the functional equation can be written as follows:

$$(1.6) \quad f_\theta(x) = g(x + \theta_0) + b^{-\alpha} f_{T\theta}(bx).$$

It is clear that the probability \mathbb{P} is T -invariant. Moreover, the 0-1 law, which is satisfied by the independent phases, states that the σ -field $B_\infty = \bigcap_{n=0}^\infty T^{-n}(B_0)$ is constituted of events with trivial probability (B_0 is the σ -field of Borel sets on Ω). Then, it is classical to conclude that the strong mixing property

$$\forall A, B \in B_0, \quad \lim_{n \rightarrow \infty} \mathbb{P}[A \cap T^{-n}B] = \mathbb{P}[A]\mathbb{P}[B]$$

is satisfied by the T -invariant probability \mathbb{P} (for elementary results on ergodic theory, see for example [13] or [19]). In particular, the probability \mathbb{P} is ergodic: invariant Borel sets have probability 0 or 1. This ergodicity will be the key point of the proof of our main theorems (Theorems 3.1 and 3.2).

Let us now briefly describe the main results of this paper. We first observe in Section 2 that a conclusion like (i) in Theorem 1.1 is not possible in the random context. More precisely, when g is not constant, f_θ is almost surely not Lipschitz. Then, we can state in Section 3 and prove in Section 4 minoration for the oscillations. More precisely, we prove that with probability one, the minoration

$$(1.7) \quad \text{osc}(f_\theta, I) \geq C |I|^{\alpha+\varepsilon}$$

is true for every interval I , as soon as g is not constant. Moreover, the stronger inequality

$$(1.8) \quad \text{osc}(f_\theta, I) \geq C |I|^\alpha$$

is valid when $|I| = \delta_n$, where $(\delta_n)_n$ is a sequence of scales decreasing to zero. In particular, the local Hölder index of f_θ is almost surely everywhere equal to α . In Section 5, we deal with the case where b is a transcendental number. In that case, we are able to prove that the minoration (1.8) is true for every interval I and for every choice of phases $(\theta_n)_{n \geq 0}$. In some sense, this means that randomness is already present in the number b .

2. WHEN g IS NOT CONSTANT, f_θ IS NOT LIPSCHITZ

The first step of our investigation is to establish that f_θ cannot be regular (except when g is constant). This is the aim of the following result.

Theorem 2.1. *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a 1-periodic Lipschitz function, $1 < b < +\infty$ and $0 < \alpha < 1$. Define f_θ by formula (1.4), $c_0 = \frac{\|g'\|_\infty}{b^{1-\alpha}-1}$, and let*

$$A = \{\theta \in \Omega ; f_\theta \text{ is Lipschitz}\} \quad \text{and} \quad A_{c_0} = \{\theta \in \Omega ; f_\theta \text{ is } c_0\text{-Lipschitz}\}.$$

Then,

$$\mathbb{P}[A] = 0 \text{ or } 1 \quad \text{and} \quad \mathbb{P}[A_{c_0}] = 0 \text{ or } 1 .$$

Moreover, the following are equivalent:

- (i) $\mathbb{P}[A] = 1$;
- (ii) $\mathbb{P}[A_{c_0}] = 1$;
- (iii) g is a constant function.

Remark 2.2. If g is not constant, Theorem 2.1 ensures that with probability one, there exist $x \neq y$ and $v > 0$ such that

$$|f_\theta(x) - f_\theta(y)| \geq \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha}-1} |x-y| .$$

In fact, the theorem also says that there is no more information once one shows that f_θ is almost surely not Lipschitz. This remark will be useful when proving Theorem 3.1.

Proof of Theorem 2.1. Observing the functional equation (1.6), it is clear that A is T -invariant. The ergodicity of the shift T then ensures that $\mathbb{P}[A] = 0$ or $\mathbb{P}[A] = 1$. More precisely, if $A_c = \{\theta \in \Omega ; f_\theta \text{ is } c\text{-Lipschitz}\}$ and $\varphi(c) = b^{\alpha-1}(c + \|g'\|_\infty)$, it is easy to check that

$$\theta \in A_c \implies T\theta \in A_{\varphi(c)} .$$

In other words,

$$(2.1) \quad A_c \subset T^{-1}(A_{\varphi(c)}) .$$

In particular, if $c \geq c_0$, then $\varphi(c) \leq c$ and $A_c \subset T^{-1}(A_c)$. We cannot state that $A_c = T^{-1}(A_c)$, but we can however conclude that $\mathbb{P}[A_c] = 0$ or $\mathbb{P}[A_c] = 1$. This well-known result is a consequence of the fact that an ergodic invariant probability measure has no wandering set of positive measure. Let us propose an elementary direct proof, in order to be self-contained. \square

Lemma 2.3. Let $B \subset \Omega$ be a Borel set such that $B \subset T^{-1}B$. Then

$$\mathbb{P}[B] = 0 \quad \text{or} \quad \mathbb{P}[B] = 1 .$$

Proof. Let $C = \Omega \setminus B$ and $D = \limsup_{n \rightarrow \infty} T^{-n}C$. Observe that D is T -invariant, and suppose that $\mathbb{P}[B] < 1$. Then $\mathbb{P}[D] > 0$ and, by ergodicity, $\mathbb{P}[D] = 1$ (here, we use the fact that \mathbb{P} is a finite measure). It follows that $\mathbb{P}[B] = \mathbb{P}[B \cap D]$. But, if $B \subset T^{-1}B$, then $B \cap D = \emptyset$. So $\mathbb{P}[B] = 0$. \square

Proof of (i) \Rightarrow (ii). Suppose that $\mathbb{P}[A] = 1$ and note that $A = \bigcup_{c \geq c_0} A_c$. So, we can find some $c \geq c_0$ such that $\mathbb{P}[A_c] > 0$. It follows that $\mathbb{P}[A_c] = 1$. Using (2.1) and the T -invariance of the measure \mathbb{P} , we obtain $\mathbb{P}[A_{\varphi(c)}] = 1$. Iterating this, we get $\mathbb{P}[A_{\varphi^n(c)}] = 1$. Taking the limit when $n \rightarrow \infty$, we conclude that $\mathbb{P}[A_{c_0}] = 1$ (c_0 is the unique fixed point of the contraction φ). \square

Proof of (ii) \Rightarrow (iii). Let x and y be two real numbers, and write

$$f_\theta(x) - f_\theta(y) = \sum_{n=0}^{+\infty} b^{-n\alpha} (g(b^n x + \theta_n) - g(b^n y + \theta_n)) .$$

The random variables $b^{-n\alpha}(g(b^n x + \theta_n) - g(b^n y + \theta_n))$ are centered and independent (here, we use the fact that g is 1-periodic). So, they are orthogonal in $L^2(\mathbb{P})$. If $p \geq 0$, the Pythagorean theorem ensures that

$$\begin{aligned} \mathbb{E} [(f_\theta(x) - f_\theta(y))^2] &= \sum_{n=0}^{+\infty} b^{-2n\alpha} \mathbb{E} [(g(b^n x + \theta_n) - g(b^n y + \theta_n))^2] \\ &\geq b^{-2p\alpha} \mathbb{E} [(g(b^p x + \theta_p) - g(b^p y + \theta_p))^2] . \end{aligned}$$

Replacing x by $b^{-p}x$ and y by $b^{-p}y$ and using (ii), we get

$$\int_0^1 (g(x+t) - g(y+t))^2 dt \leq b^{2p\alpha} c_0^2 |b^{-p}x - b^{-p}y|^2 .$$

Taking the limit when $p \rightarrow \infty$, we obtain

$$\int_0^1 (g(x+t) - g(y+t))^2 dt = 0 .$$

This implies that $g(x) = g(y)$. □

3. ON OSCILLATIONS OF f_θ

In this section, we state the main results concerning the local behavior of Weierstrass functions with random phases.

Theorem 3.1. *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a 1-periodic non-constant Lipschitz function, and let $1 < b < +\infty$ and $0 < \alpha < 1$. Define f_θ by formula (1.4). There exists a constant $C > 0$ such that for almost every $\theta \in \Omega$, we can choose a sequence of scales $\delta_k \searrow 0$ such that, for every interval I of length $|I| = \delta_k$,*

$$\text{osc}(f_\theta, I) \geq C |I|^\alpha .$$

It seems curious that we have to restrict the conclusion to some sequence of scales. In fact, it is easy to see that the more precise conclusion

$$(3.1) \quad \exists C > 0 ; \text{ for almost every } \theta \in \Omega, \forall I \text{ with } |I| \leq 1, \text{osc}(f_\theta, I) \geq C |I|^\alpha$$

is false in general. Indeed, the set of $\theta \in \Omega$ such that $\text{osc}(f_\theta, I) \geq C |I|^\alpha$ for all intervals I of length $|I| \leq 1$ is clearly a closed subset of Ω . So, conclusion (3.1) would imply the same property for every $\theta \in \Omega$. On the other hand, if b is an integer and if $g(x) = \varphi(x) - b^{-\alpha}\varphi(bx)$, with φ a 1-periodic non-constant Lipschitz function, then g is a 1-periodic non-constant Lipschitz function. Moreover, $f_\theta = \varphi$ with $\theta = (0, 0, \dots)$. So we can find some $\theta \in \Omega$ such that f_θ is regular. As we will see in section 5, this phenomenon cannot occur when b is a transcendental number.

In view of this remark, it is reasonable to ask whether the weaker assertion

$$(3.2) \quad \text{for almost every } \theta \in \Omega, \exists C > 0; \forall I \text{ with } |I| \leq 1, \text{osc}(f_\theta, I) \geq C |I|^\alpha$$

is true when g is non-constant. In fact, we do not know if (3.2) is true for every $b > 1$, but we can prove an analogue of (3.2), replacing α by $\alpha + \varepsilon$. This is the purpose of the following theorem.

Theorem 3.2. *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a 1-periodic non-constant Lipschitz function, and let $1 < b < +\infty$ and $0 < \alpha < 1$. Define f_θ by formula (1.4). For almost every $\theta \in \Omega$, the following conclusion is true:*

$$(3.3) \quad \forall \varepsilon > 0, \exists C > 0; \forall I \text{ with } |I| \leq 1, \text{osc}(f_\theta, I) \geq C |I|^{\alpha+\varepsilon} .$$

Using Theorem 3.2 and the fact that f_θ is α -Hölder, we obtain the following direct consequence about the local behavior of Weierstrass functions with phases.

Corollary 3.3. *Suppose that g is not constant. The function f_θ is almost surely nowhere differentiable. More precisely, for almost every $\theta \in \Omega$, we have*

$$\lim_{x \in I, |I| \searrow 0} \frac{\ln(\text{osc}(f_\theta, I))}{\ln(|I|)} = \alpha \quad \text{for all } x \in \mathbb{R}.$$

In particular, for almost every $\theta \in \Omega$, the Hölder index of f_θ is everywhere equal to α .

4. PROOF OF THEOREMS 3.1 AND 3.2

We begin with the following elementary lemma.

Lemma 4.1 (Transfer lemma). *Let $x, y \in \mathbb{R}$ satisfy*

$$\frac{|f_{T\theta}(y) - f_{T\theta}(x)|}{|y - x|} \geq \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha} - 1} \quad \text{for some } v > 0.$$

Then

$$\frac{|f_\theta(b^{-1}y) - f_\theta(b^{-1}x)|}{|b^{-1}y - b^{-1}x|} \geq \frac{(1+vb^{1-\alpha})\|g'\|_\infty}{b^{1-\alpha} - 1}.$$

Proof. The above lemma is an easy consequence of the functional equation (1.6). We have

$$f_\theta(b^{-1}y) - f_\theta(b^{-1}x) = g(b^{-1}y + \theta_0) - g(b^{-1}x + \theta_0) + b^{-\alpha}(f_{T\theta}(y) - f_{T\theta}(x)).$$

Suppose that x, y and v satisfy the hypothesis of the lemma. Then

$$\begin{aligned} \frac{|f_\theta(b^{-1}y) - f_\theta(b^{-1}x)|}{|b^{-1}y - b^{-1}x|} &\geq b^{1-\alpha} \frac{|f_{T\theta}(y) - f_{T\theta}(x)|}{|y - x|} - \frac{|g(b^{-1}y + \theta_0) - g(b^{-1}x + \theta_0)|}{|b^{-1}y - b^{-1}x|} \\ &\geq b^{1-\alpha} \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha} - 1} - \|g'\|_\infty \\ &= \frac{(1+vb^{1-\alpha})\|g'\|_\infty}{b^{1-\alpha} - 1}. \end{aligned}$$

□

When we apply Lemma 4.1, we shrink the scale by a factor $1/b$ and obtain a stronger estimate (the constant v is replaced by the bigger one $vb^{1-\alpha}$). Using iterations of Lemma 4.1, it is possible, from an estimate of the oscillation of an iterate $f_{T^n\theta}$ at a big scale, to get an estimate of the oscillation of f_θ at a small scale. Of course, this final minoration will make sense if we are able to control the distance $|y - x|$ when we begin to use Lemma 4.1. The following lemma is a step in this direction.

Lemma 4.2. *Let $v > 0$. If $h, \ell > 0$, denote by $M_{h,\ell}$ the set of $\theta \in \Omega$ such that for every closed interval I of length $|I| = \ell$, there exists $x \in I$ with $x + h \in I$ and*

$$|f_\theta(x + h) - f_\theta(x)| \geq \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha} - 1} |h|.$$

Then there exist $h_0 > 0$ and $\ell_0 \geq 1$ such that

$$\mathbb{P}[M_{h_0, \ell_0}] > 0.$$

Proof. Let $v > 0$ and suppose that g is not constant (this is the only interesting case). According to Theorem 2.1, the set

$$M = \left\{ \theta \in \Omega; \exists x \in \mathbb{R}, \exists h > 0; |f_\theta(x+h) - f_\theta(x)| \geq \frac{(1+3v)\|g'\|_\infty}{b^{1-\alpha} - 1} |h| \right\}$$

is such that $\mathbb{P}[M] = 1$. We can then find $h_0 > 0$ such that the set M_{h_0} of θ satisfying

$$(4.1) \quad |f_\theta(x_\theta + h_0) - f_\theta(x_\theta)| \geq \frac{(1+3v)\|g'\|_\infty}{b^{1-\alpha} - 1} |h_0| \quad \text{for some } x_\theta \in \mathbb{R}$$

has positive probability. Let

$$\varepsilon = \frac{v\|g'\|_\infty}{b^{1-\alpha} - 1} |h_0|.$$

According to Proposition 1.2, we can find a real number $\ell_0 > 0$ (which can be supposed greater than 1 and greater than $2h_0$) such that

$$\forall \theta \in \Omega, \forall \delta \in \mathbb{R}, \exists a \in [\delta, \delta + \ell_0/2); \|\tau_a f_\theta - f_\theta\|_\infty \leq \varepsilon.$$

Let $I = [z, z + \ell_0]$ be a closed interval of length ℓ_0 . Suppose that θ satisfies (4.1) and take $\delta = z - x_\theta$. We can find $a \in [\delta, \delta + \ell_0/2)$ such that

$$\|\tau_a f_\theta - f_\theta\|_\infty \leq \varepsilon.$$

If $\tilde{x}_\theta = a + x_\theta$, we have

$$z \leq \tilde{x}_\theta < \tilde{x}_\theta + h_0 \leq z + \ell_0.$$

Moreover,

$$\begin{cases} |f_\theta(\tilde{x}_\theta) - f_\theta(x_\theta)| = |\tau_a f_\theta(x_\theta) - f_\theta(x_\theta)| \leq \varepsilon, \\ |f_\theta(\tilde{x}_\theta + h_0) - f_\theta(x_\theta + h_0)| = |\tau_a f_\theta(x_\theta + h_0) - f_\theta(x_\theta + h_0)| \leq \varepsilon, \end{cases}$$

and it follows that

$$\frac{|f_\theta(\tilde{x}_\theta + h_0) - f_\theta(\tilde{x}_\theta)|}{|h_0|} \geq \frac{|f_\theta(x_\theta + h_0) - f_\theta(x_\theta)| - 2\varepsilon}{|h_0|} \geq \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha} - 1}.$$

We have just proved that $M_{h_0} \subset M_{h_0, \ell_0}$, and we can conclude that

$$\mathbb{P}[M_{h_0, \ell_0}] \geq \mathbb{P}[M_{h_0}] > 0.$$

□

We can now finish the proof of Theorem 3.1. Suppose that g is not constant and let $v > 0$. According to Lemma 4.2, choose $h_0 > 0$ and $\ell_0 \geq 1$ such that

$$\mathbb{P}[M_{h_0, \ell_0}] > 0.$$

Using the ergodicity of the probability measure \mathbb{P} , we can claim that for almost every $\theta \in \Omega$, $T^n(\theta) \in M_{h_0, \ell_0}$ infinitely often (the set $\limsup_{n \rightarrow \infty} T^{-n}(M_{h_0, \ell_0})$ is T -invariant with positive probability).

Denote by E this set of full measure. If $\theta \in E$, we can construct a sequence of integers $(n_k)_{k \geq 0}$ increasing to $+\infty$ such that

$$T^{n_k}(\theta) \in M_{h_0, \ell_0} \quad \text{for all } k \geq 0.$$

Let $\delta_k = \ell_0 b^{-n_k}$. If I is an interval of length $|I| = \delta_k$ and if $\tilde{I} = b^{n_k} I$, we can find $x \in \tilde{I}$ such that $x + h_0 \in \tilde{I}$ and

$$|f_{T^{n_k} \theta}(x + h_0) - f_{T^{n_k} \theta}(x)| \geq \frac{(1+v)\|g'\|_\infty}{b^{1-\alpha} - 1} |h_0|.$$

Iterating Lemma 4.1, we obtain

$$|f_\theta(b^{-n_k}(x+h_0)) - f_\theta(b^{-n_k}x)| \geq \frac{(1+vb^{n_k(1-\alpha)})\|g'\|_\infty}{b^{1-\alpha}-1} |b^{-n_k}h_0|.$$

Take

$$C = \frac{v\|g'\|_\infty|h_0|}{\ell_0^\alpha(b^{1-\alpha}-1)}.$$

This constant does not depend on θ and k and is such that

$$\text{osc}(f_\theta, I) \geq |f_\theta(b^{-n_k}(x+h_0)) - f_\theta(b^{-n_k}x)| \geq \frac{vb^{n_k(1-\alpha)}\|g'\|_\infty}{b^{1-\alpha}-1} |b^{-n_k}h_0| = C|I|^\alpha.$$

This completes the proof of Theorem 3.1. \square

Let us now prove Theorem 3.2. As previously, we choose $v > 0$, $h_0 > 0$ and $\ell_0 \geq 1$ such that

$$\mathbb{P}[M_{h_0, \ell_0}] > 0.$$

We need more quantitative information about the sequence of iterates T^{n_k} satisfying $T^{n_k}(\theta) \in M_{h_0, \ell_0}$. This will be given by the ergodic theorem, which gives us the asymptotic behavior of the frequency of the returns in M_{h_0, ℓ_0} (for a simple proof of the ergodic theorem, see for example [4], page 98). More precisely, this theorem states that almost surely,

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{M_{h_0, \ell_0}} \circ T^j \xrightarrow{n \rightarrow \infty} \mathbb{P}[M_{h_0, \ell_0}].$$

In other words,

$$(4.2) \quad \lim_{k \rightarrow \infty} \frac{n_k(\theta)}{k} = \frac{1}{\mathbb{P}[M_{h_0, \ell_0}]} \quad \text{almost surely,}$$

where $n_k(\theta)$ is the k -th return in M_{h_0, ℓ_0} .

Denote by \tilde{E} the set of θ satisfying (4.2), and let $\varepsilon > 0$. If $\theta \in \tilde{E}$, the sequence $n_k(\theta)/n_{k-1}(\theta)$ tends to 1 at infinity. So, we can find $k_0 \geq 1$ such that

$$(4.3) \quad \forall k \geq k_0, \quad \frac{n_k(\theta)}{n_{k-1}(\theta)} \leq \frac{\alpha + \varepsilon}{\alpha}.$$

For simplicity, write n_k instead of $n_k(\theta)$ and let I be an interval of length $|I| \leq \ell_0 b^{-n_{k_0-1}}$. There exists $k \geq k_0$ such that

$$\ell_0 b^{-n_k} \leq |I| \leq \ell_0 b^{-n_{k-1}}.$$

We can then choose an interval $J \subset I$ with length $|J| = \ell_0 b^{-n_k}$. Using Lemma 4.1 as in the proof of Theorem 3.1, we get

$$\text{osc}(f_\theta, I) \geq \text{osc}(f_\theta, J) \geq \frac{(1+vb^{n_k(1-\alpha)})\|g'\|_\infty}{b^{1-\alpha}-1} |b^{-n_k}h_0| \geq \frac{v|h_0|\|g'\|_\infty}{b^{1-\alpha}-1} |b^{-\alpha n_k}|.$$

The choice of k_0 ensures that

$$b^{-\alpha n_k} \geq \frac{1}{\ell_0^{\alpha+\varepsilon}} |I|^{\alpha+\varepsilon}.$$

The conclusion follows for every interval I of length $|I| \leq \ell_0 b^{-n_{k_0-1}}$ if we put

$$C = \frac{v|h_0|\|g'\|_\infty}{\ell_0^{\alpha+\varepsilon}(b^{1-\alpha}-1)}.$$

Finally, replacing C by $\min(C, C(\ell_0 b^{-n_{k_0-1}})^{\alpha+\varepsilon})$, we have the conclusion for any interval I of length $|I| \leq 1$. \square

Final remark on Theorem 3.2. With the method used in the proof of Theorem 3.2, we would have to know that $n_{k+1}(\theta) - n_k(\theta)$ is almost surely bounded in order to conclude that (3.3) is true when $\varepsilon = 0$. Of course, this assertion is false: the number of iterates between two consecutive returns in M_{h_0, ℓ_0} cannot be uniformly bounded, and the ergodic theorem only ensures that $n_{k+1}(\theta)/n_k(\theta)$ tends to 1 with probability one.

5. THE CASE WHERE b IS TRANSCENDENTAL

In the case where b is a transcendental number, we can improve the conclusion of Theorems 3.1 and 3.2. This is the purpose of the following result.

Theorem 5.1. *Let $b > 1$ be a transcendental number, $0 < \alpha < 1$, and $g : \mathbb{R} \rightarrow \mathbb{R}$ a 1-periodic non-constant Lipschitz function. Define f_θ by formula (1.4). There exists a constant $C > 0$ such that for every $\theta \in \Omega$ and for every interval I of length $|I| \leq 1$,*

$$\text{osc}(f_\theta, I) \geq C|I|^\alpha.$$

Remark 5.2. In particular, the conclusion is true when $\theta = (0, 0, \dots)$. This seems to be in contradiction with the possible case (i) in Theorem 1.1. In fact, it means that if f is almost periodic, Lipschitz and non-constant, and if b is a transcendental number, then $g(x) = f(x) - b^{-\alpha}f(bx)$ cannot be 1-periodic. An elementary proof of this property can be obtained by Fourier analysis. For algebraic numbers, such a construction is possible. For example, if $b = \sqrt{2}$ and if $g(x) = \cos(2\pi x) - 2^{-\alpha} \cos(4\pi x)$, we obtain

$$f(x) = \sum_{n=0}^{+\infty} b^{-n\alpha} g(b^n x) = \cos(2\pi x) + 2^{-\alpha/2} \cos(2^{3/2}\pi x),$$

which is a regular function.

Proof of Theorem 5.1. The proof of Theorem 5.1 is a consequence of the following lemma, which replaces Lemma 4.2. \square

Lemma 5.3. *In the notation of Lemma 4.2, suppose that b is a transcendental number. Then, for all $v > 0$, there exist $h_0 > 0$ and $\ell_0 > 0$ such that*

$$M_{h_0, \ell_0} = \Omega.$$

Proof. Remember the notation in the proof of Lemma 4.2, and let h_0 be a positive number such that $\mathbb{P}[M_{h_0}] > 0$. Let $t \neq 0$ and define the translation $S : \Omega \rightarrow \Omega$ by

$$(S\theta)_n = \theta_n + tb^n \pmod{1}.$$

We know that in the compact group $(\mathbb{R}/\mathbb{Z})^{\mathbb{N}}$ (equipped with the Haar measure), the translation by an element $\lambda = (\lambda_n)_{n \geq 0}$ is ergodic if and only if the real numbers $1, \lambda_0, \dots, \lambda_n, \dots$ are \mathbb{Q} -independent (it suffices to note that invariant L^2 functions f are exactly those whose Fourier coefficients satisfy $\hat{f}(n_0, \dots, n_p) = e^{2i\pi(n_0\lambda_0 + \dots + n_p\lambda_p)} \hat{f}(n_0, \dots, n_p)$ for every $p \geq 0$). When b is a transcendental number, we can then conclude that S is \mathbb{P} -invariant and ergodic. On the other hand, as an easy consequence of the relation $f_{S\theta}(x - t) = f_\theta(x)$, we note that M_{h_0} is invariant under the translation S . It follows that $\mathbb{P}[M_{h_0}] = 1$. If ℓ_0 is constructed as in Lemma 4.2, we get $\mathbb{P}[M_{h_0, \ell_0}] = 1$. In particular, M_{h_0, ℓ_0} is dense in Ω . Moreover, if a sequence $(\theta^n)_{n \geq 0}$ of points of M_{h_0, ℓ_0} converges to $\theta \in \Omega$, we know that

the sequence f_{θ^n} converges uniformly to f_θ . Using the compactness of the intervals $I = [z, z + \ell_0]$ of length ℓ_0 , it is then easy to conclude that $\theta \in M_{h_0, \ell_0}$. Finally, M_{h_0, ℓ_0} is dense and closed in Ω . This means that $M_{h_0, \ell_0} = \Omega$. \square

Remark 5.4. Let $B = (b^n)_{n \geq 0}$ with b a transcendental number. For every $\theta \in \Omega$, the projection of the curve $(\theta + tB)_{t \in \mathbb{R}}$ in the infinite-dimensional torus $(\mathbb{R}/\mathbb{Z})^{\mathbb{N}}$ is dense in $(\mathbb{R}/\mathbb{Z})^{\mathbb{N}}$. If we can prove that $M_{h_0} \neq \emptyset$ for some $h_0 > 0$, we can conclude that M_{h_0} is dense in Ω and then obtain another proof of Lemma 5.3. In fact, it is easy to construct $\theta \in \Omega$ and $h_0 > 0$ such that $\theta \in M_{h_0}$. Let $x_0 \in [0, 1]$ be a point where g is minimal, and let $y_0 > 0$ be such that $g(y_0 + x_0) > g(x_0)$. Let $\theta = (x_0, x_0, \dots, x_0, \dots)$. We have

$$\begin{aligned} f_\theta(b^{-p}y_0) - f_\theta(0) &= \sum_{n=0}^{+\infty} b^{-n\alpha} [g(b^{n-p}y_0 + x_0) - g(x_0)] \\ &\geq b^{-p\alpha} [g(y_0 + x_0) - g(x_0)] \\ &\geq \frac{(1 + 3v)\|g'\|_\infty}{b^{1-\alpha} - 1} |b^{-p}y_0| \quad \text{if } p \text{ is sufficiently large,} \end{aligned}$$

and we can conclude that $\theta \in M_{b^{-p}y_0}$.

Let us now sketch the end of the proof of Theorem 5.1. The main argument is the one used for the previous proofs. Let $\theta \in \Omega$, and let I be an interval such that $\ell_0 b^{-k} \leq |I| \leq \ell_0 b^{-k+1}$. An iteration of Lemma 5.3 gives

$$\text{osc}(f_\theta, I) \geq \text{osc}(f_\theta, J) \geq \frac{(1 + vb^{k(1-\alpha)})\|g'\|_\infty}{b^{1-\alpha} - 1} |b^{-k}h_0| \geq \frac{v|h_0|\|g'\|_\infty}{b^{1-\alpha} - 1} \left[\frac{|I|}{\ell_0 b} \right]^\alpha,$$

where $J \subset I$ has length $\ell_0 b^{-k}$. We can choose $C = \frac{v|h_0|\|g'\|_\infty}{(b^{1-\alpha}-1)(\ell_0 b)^\alpha}$.

6. ON FRACTAL DIMENSION OF THE GRAPH OF f_θ

Theorems 3.2 and 5.1 allow us to estimate the box-counting dimension and the packing dimension of the graph of f_θ .

Theorem 6.1. *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a 1-periodic non-constant Lipschitz function, and let $1 < b < +\infty$ and $0 < \alpha < 1$. Define f_θ by formula (1.4) and $\Gamma_\theta(A)$ by*

$$\Gamma_\theta(A) = \{(x, f_\theta(x)) \mid x \in A\}.$$

We have

(i) *for every non-trivial compact interval I , $\Delta(\Gamma_\theta(I)) = 2 - \alpha$ almost surely and*

(ii) *for every non-trivial interval I , $\text{Dim}(\Gamma_\theta(I)) = 2 - \alpha$ almost surely, where $\Delta(E)$ and $\text{Dim}(E)$ are respectively the box-counting dimension and the packing dimension of a set E .*

Moreover, when b is a transcendental number, conclusions (i) and (ii) are valid for every $\theta \in \Omega$.

Statement (i) is well known if we have a minoration of the oscillations (see, for example, [3]). The minoration of the packing dimension is probably less known (for general information about the packing dimension, see [18] or [17]). We can state the following elementary proposition.

Proposition 6.2. *Let $0 < \gamma < 1$, and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a function such that*

$$\text{osc}(h, I) \geq C|I|^\gamma \quad \text{for all intervals } I \text{ of length } |I| \leq 1.$$

Then, for every interval I with nonempty interior,

$$\text{Dim}(\Gamma(I)) \geq 2 - \gamma,$$

where $\Gamma(A) = \{(x, h(x)) \mid x \in A\}$.

Proof. We can suppose that $I = [a, b]$ is a compact interval with $a \neq b$ (the quantity Dim is σ -stable). Remember that if $E \subset \mathbb{R}^2$, then

$$\text{Dim}(E) = \inf \left\{ \sup \Delta(E_i); E \subset \bigcup_{i=1}^{\infty} E_i \text{ where the } E_i \text{ are closed sets} \right\}.$$

Suppose that $\Gamma([a, b]) \subset \bigcup_{i=1}^{\infty} E_i$. We can then find a sequence (F_i) of closed subsets of $[a, b]$ with

$$[a, b] = \bigcup_{i=1}^{\infty} F_i \quad \text{and} \quad \forall i \geq 1, \Gamma(F_i) \subset E_i.$$

Baire's property ensures that we can find i_0 such that the interior of F_{i_0} is not empty. It follows that $\Delta(E_{i_0}) \geq \Delta(\Gamma(F_{i_0})) \geq 2 - \gamma$, and we can conclude that $\text{Dim}(\Gamma([a, b])) \geq 2 - \gamma$. \square

ACKNOWLEDGMENTS

The author wants to thank Claude Tricot for helpful discussions about Weierstrass functions with phases. Thanks also to Guillaume Havard for his interest in this work and for his careful reading of the first drafts.

REFERENCES

1. T. Bousch and Y. Heurteaux, *On oscillations of Weierstrass-type functions*, manuscript, 1999.
2. ———, *Caloric measure on domains bounded by Weierstrass-type graphs*, Ann. Acad. Sci. Fenn. **25** (2000), 501–522. MR **2001h**:31004
3. K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, New York, 1990. MR **92j**:28008
4. ———, *Techniques in fractal geometry*, John Wiley & Sons, New York, 1997. MR **99f**:28013
5. G. H. Hardy, *Weierstrass's non-differentiable function*, Trans. Amer. Math. Soc. **17** (1916), 301–325.
6. T.-Y. Hu and K.-S. Lau, *The sum of Rademacher functions and Hausdorff dimension*, Math. Proc. Cambridge Philos. Soc. **108** (1990), 97–103. MR **91d**:28020
7. ———, *Fractal dimensions and singularities of the Weierstrass type functions*, Trans. Amer. Math. Soc. **335** (1993), 649–665. MR **93d**:28011
8. B. R. Hunt, *The Hausdorff dimension of graphs of Weierstrass functions*, Proc. Amer. Math. Soc. **126** (1998), 791–800. MR **98i**:28009
9. J. L. Kaplan, J. Mallet-Paret, and J. A. Yorke, *The Lyapunov dimension of a nowhere differentiable attracting torus*, Ergodic Theory & Dynamical Systems **4** (1984), 261–281. MR **86h**:58091
10. F. Ledrappier, *On the dimension of some graphs*, Contemp. Math. **135** (1992), 285–293. MR **94d**:28007
11. R. D. Mauldin and S. C. Williams, *On the Hausdorff dimension of some graphs*, Trans. Amer. Math. Soc. **298** (1986), 793–804. MR **88c**:28006
12. C. McMullen, *The Hausdorff dimension of general Sierpinski carpets*, Nagoya Math. J. **96** (1984), 1–9. MR **86h**:11061
13. K. Petersen, *Ergodic theory*, Cambridge University Press, Cambridge, 1983. MR **87i**:28002
14. F. Przytycki and M. Urbanski, *On the Hausdorff dimension of some fractal sets*, Studia Math. **93** (1989), 155–186. MR **90f**:28006

15. Y. Shiota and T. Sekiguchi, *Hausdorff dimension of graphs of some Rademacher series*, Japan J. Appl. Math. **7** (1990), 121–129. MR **91e**:28009
16. J. Szulga, *Hausdorff dimension of Weierstrass-Mandelbrot process*, Statist. Probab. Lett. **56** (2002), 301–307. MR **2002m**:60069
17. C. Tricot, *Sur la classification des ensembles boréliens de mesure de Lebesgue nulle*, Ph.D. thesis, Faculté des Sciences de l'Université de Genève, 1980.
18. ———, *Two definitions of fractional dimension*, Math. Proc. Cambridge Philos. Soc. **91** (1982), 57–74. MR **84d**:28013
19. P. Walters, *An introduction to ergodic theory*, Springer-Verlag, New York, 1982. MR **84e**:28017

LABORATOIRE DE MATHÉMATIQUES PURES, UNIVERSITÉ BLAISE PASCAL, F-63177 AUBIÈRE
CEDEX, FRANCE

E-mail address: Yanick.Heurteaux@math.univ-bpclermont.fr

EXPLICIT LOWER BOUNDS FOR RESIDUES AT $s = 1$ OF DEDEKIND ZETA FUNCTIONS AND RELATIVE CLASS NUMBERS OF CM-FIELDS

STÉPHANE LOUBOUTIN

Dedicated to Jacqueline G.

ABSTRACT. Let S be a given set of positive rational primes. Assume that the value of the Dedekind zeta function ζ_K of a number field K is less than or equal to zero at some real point β in the range $\frac{1}{2} < \beta < 1$. We give explicit lower bounds on the residue at $s = 1$ of this Dedekind zeta function which depend on β , the absolute value d_K of the discriminant of K and the behavior in K of the rational primes $p \in S$. Now, let k be a real abelian number field and let β be any real zero of the zeta function of k . We give an upper bound on the residue at $s = 1$ of ζ_k which depends on β , d_k and the behavior in k of the rational primes $p \in S$. By combining these two results, we obtain lower bounds for the relative class numbers of some normal CM-fields K which depend on the behavior in K of the rational primes $p \in S$. We will then show that these new lower bounds for relative class numbers are of paramount importance for solving, for example, the exponent-two class group problem for the non-normal quartic CM-fields. Finally, we will prove Brauer-Siegel-like results about the asymptotic behavior of relative class numbers of CM-fields.

The main results arrived at in this paper are Theorems 1, 14, 22 and 26.

1. LOWER BOUNDS FOR RESIDUES OF ZETA FUNCTIONS

Let $c > 0$ be given (to be selected below). It has long been known that Hecke's integral representations of Dedekind zeta functions ζ_K of number fields K can be used to obtain lower bounds for their residues κ_K at $s = 1$ of the type

$$1 - (c/\log d_K) \leq \beta < 1 \text{ and } \zeta_K(\beta) \leq 0 \text{ imply } \kappa_K \geq (1 - \beta)d_K^{(\beta-1)/2}(1 + o(1)),$$

where $o(1)$ is an error term that approaches zero as $d_K \rightarrow \infty$ provided that K ranges over number fields of a given degree (e.g. see [Lou2, Proposition A]. See also [Lan, Chapter XVI, Section 2, Lemma 3, p. 323] for a weaker result). Notice that the best lower bound one can deduce (for $\beta = 1 - (2/\log d_K)$) is of the type

$$\zeta_K(1 - (2/\log d_K)) \leq 0 \text{ implies } \kappa_K \geq \frac{2}{e \log d_K}(1 + o(1)).$$

The first aim of this paper is to prove Theorem 1 below, which not only provides a nice treatment of this error term (by simply getting rid of it!) but also allows us to obtain lower bounds for these residues which depend on the behavior in K of a

Received by the editors April 23, 2002 and, in revised form, January 6, 2003.

2000 *Mathematics Subject Classification.* Primary 11R42; Secondary 11R29.

Key words and phrases. Dedekind zeta functions, CM-field, relative class number.

finite set S of rational primes. Let us first set some notation. If K is an algebraic number field and S is any finite set of positive rational primes, we define

$$\Pi_K(S) := \prod_{p \in S} \prod_{\mathcal{P}|p} (1 - (N(\mathcal{P}))^{-1})^{-1} \geq 1$$

(product of Euler's factors of the Dedekind zeta function of K) and

$$\Lambda_S := \prod_{p \in S} (1 + p^{-1/2})^4 \geq 1,$$

with the convention $\Pi_K(\emptyset) = \Lambda_\emptyset = 1$. Our first result is as follows:

Theorem 1.

(1) Let $m \geq 1$ be a positive integer. There exists ρ_{2m} effective such that for any finite set S of primes and any totally imaginary number field K of degree $2n \geq 2m$ and root discriminant $\rho_K := d_K^{1/2n} \geq \rho_{2m} \Lambda_S$ we have

$$(1) \quad \kappa_K \geq \frac{1}{2}(1 - \beta)d_K^{(\beta-1)/2} \Pi_K(S)$$

if $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1)$.

(2) Let $m \geq 1$ be a positive integer. Let S be any given finite set of primes. There exists $\rho_{2m,S}$ effective such that for any totally imaginary number field K of degree $2n \geq 2m$ and root discriminant $\rho_K := d_K^{1/2n} \geq \rho_{2m,S}$ we have

$$(2) \quad \kappa_K \geq (1 - \beta)d_K^{(\beta-1)/2} \Pi_K(S)$$

if $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1 - (1/\log d_K)]$.

(3) We may take $\rho_{12} = \rho_{12,\emptyset} = 2\pi^2$ and for smaller values of m we may take ρ_{2m} and $\rho_{2m,S}$ for $S = \emptyset$ and $S = \{2\}$ as given in Table 1:

Table 1

$2n \geq 2m =$	2	4	6	8	10	12	∞
$\rho_{2m} =$	270	41	26	22	21	$2\pi^2$	$2\pi^2$
$\rho_{2m,\emptyset} =$	2600	50	25	20	$2\pi^2$	$2\pi^2$	$2\pi^2$
$\rho_{2m,\{2\}} =$	36000	650	295	222	194	181	$2\pi^2 \Lambda_{\{2\}} = 167.63 \dots$

(4) Let K be a totally imaginary number field of degree $2n > 2$ and root discriminant $\rho_K \geq 32\pi^2 \Lambda_{\{2\}} = 2682.208 \dots$. Assume that $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1)$. Then,

$$(3) \quad \kappa_K \geq (1 - \beta)d_K^{(\beta-1)/2}.$$

Proof. See Section 2 below. □

We could have stated this result in the more general setting of the not necessarily totally imaginary number fields. However, we only aim at using it for obtaining good lower bounds for relative class numbers of CM-fields. Notice that, contrary to our previous lower bounds given in [Lou2, Proposition A], our present lower bounds (1) and (2) do not depend on any pesky error factor

$$\epsilon_K = \max\left(1 - (2\pi n/\rho_K^\beta), \frac{2}{5} \exp(-2\pi n/\rho_K)\right),$$

which for a given n approaches 1 as $d_K \rightarrow \infty$, but which approaches 0 as $n \rightarrow \infty$ as K ranges over CM-fields of bounded root discriminants. Moreover, the real draw of these lower bounds (1) and (2) is that the Euler factors $\Pi_K(S)$ being always greater

than or equal to one, these bounds can be considerably better than the ones without the factor $\Pi_K(S)$ given in [Lou2, Proposition A]. For example, if $S = \{2\}$ and 2 splits completely in K , then $\Pi_K(S) = 4^n$. We also refer the reader to [Hof, Lemma 4] and [Sta3, Lemma 4] where other similar but less satisfactory lower bounds for κ_K are proved (in the case that $S = \emptyset$).

2. PROOF OF THEOREM 1

Let K be a totally imaginary number field of degree $2n \geq 2$. Let $\zeta_K(s)$ and d_K be the the Dedekind zeta function and the absolute value of the discriminant of K , and set $A_K = \sqrt{d_K}/(2\pi)^{2n} = (\rho_K/2\pi)^n$, $F_K(s) = A_K^s \Gamma^n(s) \zeta_K(s)$ and $\lambda_K = \text{Res}_{s=1}(F_K) = A_K \kappa_K$. Let

$$H_n(x) = \frac{1}{2\pi i} \int_{\Re(z)=\alpha} \Gamma^n(z) x^{-z} dz \quad (\alpha > 1 \text{ and } x > 0)$$

be the inverse Mellin transform of $\Gamma^n(s)$. Hence, $H_n(x) > 0$ for $x > 0$. Let $S_K(x)$ be the inverse Mellin transform of $F_K(s)$. For $x > 0$ we have

$$S_K(x) = \frac{1}{2\pi i} \int_{\Re(z)=\alpha} F_K(z) x^{-z} dz = \sum_{\mathcal{I}} H_n(xN(\mathcal{I})/A_K)$$

(where \mathcal{I} ranges over the nonzero integral ideals of K). Now, by shifting the vertical line of integration $\Re(z) = \alpha > 1$ to the left to the vertical line $\Re(z) = 1 - \alpha < 0$, by using the functional equation $F_K(1 - z) = F_K(z)$ to come back to the vertical line of integration $\Re(z) = \alpha$ and by noticing that we pick up only two poles, a simple pole of residue λ_K at $z = 1$ and a simple pole of residue $-\lambda_K$ at $z = 0$, we obtain that $S_K(x)$ satisfies the following functional equation:

$$S_K(1/x) = \lambda_K x - \lambda_K + x S_K(x).$$

Using this functional equation and the fact that $F_K(s)$ is the Mellin transform of $S_K(x)$, we obtain:

$$\begin{aligned} F_K(s) &= \int_0^\infty S_K(x) x^s \frac{dx}{x} = \int_1^\infty S_K(1/x) x^{-s} \frac{dx}{x} + \int_1^\infty S_K(x) x^s \frac{dx}{x} \\ &= \frac{\lambda_K}{s(s-1)} + \int_1^\infty S_K(x) (x^s + x^{1-s}) \frac{dx}{x} \end{aligned}$$

and

$$F_K(s) = \frac{\lambda_K}{s(s-1)} + \sum_{\mathcal{I}} \int_1^\infty H_n(xN(\mathcal{I})/A_K) (x^s + x^{1-s}) \frac{dx}{x}$$

(where \mathcal{I} ranges over the nonzero integral ideals of K), which is nothing but the Hecke integral representation of $\zeta_K(s)$, in another guise (see [Lan, Chapter XIII, Section 3, Theorem 3, p. 260]). Let S be a finite set of distinct rational primes. Set $\mathcal{S} = \{\mathcal{I}; p \mid N(\mathcal{I}) \Rightarrow p \in S\}$ and

$$\zeta_S(s) = \sum_{\mathcal{I} \in \mathcal{S}} (N(\mathcal{I}))^{-s} = \prod_{p \in S} \prod_{\mathcal{P} \mid p} (1 - (N(\mathcal{P}))^{-s})^{-1}$$

(hence, $\zeta_S(1) = \Pi_K(S)$). Since $H_n(x) > 0$ for $x > 0$, for $1 - \alpha < s < \alpha$ we have

$$\begin{aligned} F_K(s) &\geq \frac{\lambda_K}{s(s-1)} + \sum_{\mathcal{I} \in \mathcal{S}} \int_1^\infty H_n(xN(\mathcal{I})/A_K)(x^s + x^{1-s}) \frac{dx}{x} \\ &= \frac{\lambda_K}{s(s-1)} + \sum_{\mathcal{I} \in \mathcal{S}} \frac{1}{2\pi i} \int_{\Re(z)=\alpha} \Gamma^n(z) (A_K/N(\mathcal{I}))^z \left(\frac{1}{z-s} + \frac{1}{z-(1-s)} \right) dz \\ &= \frac{\lambda_K}{s(s-1)} + \frac{1}{2\pi i} \int_{\Re(z)=\alpha} \Gamma^n(z) A_K^z \zeta_S(z) \left(\frac{1}{z-s} + \frac{1}{z-(1-s)} \right) dz. \end{aligned}$$

Since $0 < s < 1$ and $\zeta_K(s) \leq 0$ imply $F_K(s) \leq 0$, we obtain:

Lemma 2. Fix $\alpha > 1$. Assume that $\zeta_K(\beta) \leq 0$ for some $\beta \in (0, 1)$. Then,

$$\kappa_K \geq \frac{\beta(1-\beta)}{2\pi i} \int_{\Re(z)=\alpha} \Gamma^n(z) A_K^{z-1} \zeta_S(z) \left(\frac{1}{z-\beta} + \frac{1}{z-(1-\beta)} \right) dz.$$

From now on, we assume that $\frac{1}{2} < \beta < 1$. We set

$$f_n(\beta) = \beta(2\pi)^{n(1-\beta)} \Gamma^n(\beta),$$

$$M(\beta) = \sup_{\Re(s)=1/2} \left| \frac{1}{s-\beta} + \frac{1}{s-(1-\beta)} \right| = \sup_{-\infty < t < \infty} \frac{2|t|}{(\beta - \frac{1}{2})^2 + t^2} = \frac{2}{2\beta-1}$$

and

$$I_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Gamma(\frac{1}{2} + it)|^n dt = \frac{1}{\pi} \int_0^{\infty} \left(\frac{\pi}{\cosh(\pi t)} \right)^{n/2} dt = \pi^{\frac{n}{2}-2} J_n,$$

where

$$J_n = \int_0^{\pi/2} \sin^{n/2-1}(T) dT$$

decreases towards zero as $n \geq 1$ goes to infinity (use $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$ to obtain $|\Gamma(1/2 + it)|^2 = \pi/\cosh(\pi t)$, and set $\cosh(\pi t) = 1/\sin T$). Notice that $J_{n+4} = \frac{n}{n+2} J_n$ for $n \geq 1$ and that $J_2 = \pi/2$ and $J_4 = 1$.

In Lemma 2, we shift the vertical line of integration $\Re(z) = \alpha > 1$ leftwards to the vertical line $\Re(z) = 1/2$. We pick up only one residue, at $z = \beta$, and obtain:

$$\begin{aligned} \kappa_K &\geq \beta(1-\beta) \left(\Gamma^n(\beta) A_K^{\beta-1} \zeta_S(\beta) - M(\beta) \zeta_S(1/2) A_K^{-1/2} I_n \right) \\ &\geq (1-\beta) d_K^{(\beta-1)/2} \zeta_S(1) \left(f_n(\beta) - \frac{2\beta d_K^{(1-\beta)/2}}{2\beta-1} \frac{\zeta_S(1/2)}{\zeta_S(1)} \left(\frac{2\pi^2}{\rho_K} \right)^{n/2} I_n \right) \\ &\geq (1-\beta) d_K^{(\beta-1)/2} \Pi_K(S) \left(f_n(\beta) - \frac{2\beta d_K^{(1-\beta)/2} J_n}{(2\beta-1)\pi^2} \left(\frac{2\pi^2 \Lambda_S}{\rho_K} \right)^{n/2} \right), \end{aligned}$$

for

$$\frac{\zeta_S(1/2)}{\zeta_S(1)} = \prod_{\mathcal{P}|p} (1 + (N(\mathcal{P}))^{-1/2}) \leq (1 + p^{-1/2})^{2n} = \Lambda_S^{n/2}$$

and $\zeta_S(1) = \Pi_K(S)$.

Lemma 3. Let $\gamma = 0.577215 \dots$ denote Euler's constant and set $f_n(\beta) := \beta(2\pi)^{n(1-\beta)}\Gamma^n(\beta)$. In the range $0 < \beta \leq 1$, it follows that

$$f_n(\beta) \geq 1 - (1 - \beta)f'_n(1) = 1 + n(1 - \beta)(\gamma + \log(2\pi) - \frac{1}{n}) \geq 1.$$

Proof. Since $f_n(\beta)$ is positive and log-convex in the range $\beta > 0$ (use the infinite product of the Γ -function), f_n is convex in the same range. \square

Using Lemma 3, noticing that $1/2 < 1 - (2/\log d_K) \leq \beta < 1$ implies $\beta/(2\beta - 1) \leq (n \log \rho_K - 1)/(n \log \rho_K - 2)$ and $d_K^{(1-\beta)/2} \leq e$, and noticing that $1 - (2/\log d_K) \leq \beta \leq 1 - (1/\log d_K)$ implies $d_K^{(1-\beta)/2}/(1 - \beta) \leq \sqrt{e} \log d_K$, we finally obtain:

Proposition 4. Let K be a totally imaginary number field of degree $2n \geq 2$, and assume that $d_K \geq e^4$.

(1) Assume that $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1)$. Then,

$$(4) \quad \kappa_K \geq (1 - \beta)d_K^{(\beta-1)/2} \Pi_K(S) \left(1 - \frac{n \log \rho_K - 1}{n \log \rho_K - 2} \frac{2eJ_n}{\pi^2} \left(\frac{2\pi^2 \Lambda_S}{\rho_K} \right)^{n/2} \right).$$

(2) Assume that $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1 - (1/\log d_K)]$. Then,

$$(5) \quad \kappa_K \geq (1 - \beta)d_K^{(\beta-1)/2} \Pi_K(S) \left(1 + n(1 - \beta)S_K \right)$$

where

$$S_K = \gamma + \log(2\pi) - \frac{1}{n} - \frac{n \log \rho_K - 1}{n \log \rho_K - 2} \frac{4\sqrt{e}J_n}{\pi^2} R_K$$

and where

$$R_K = \left(\frac{2\pi^2 \Lambda_S}{\rho_K} \right)^{n/2} \log \rho_K$$

decreases with ρ_K in the range $\rho_K \geq e^{2/n}$, i.e., in the range $d_K \geq e^4$.

Now, we are in a position to complete the proof of Theorem 1.

(1) Assume that $n \geq 6$ and $\rho_K \geq 2\pi^2 \Lambda_S \geq 2\pi^2$. We have $J_n \leq J_6 = \pi/4$, $(n \log \rho_K - 1)/(n \log \rho_K - 2) \leq (6 \log(2\pi^2) - 1)/(6 \log(2\pi^2) - 2)$, and we obtain

$$\frac{n \log \rho_K - 1}{n \log \rho_K - 2} \frac{2eJ_n}{\pi^2} \left(\frac{2\pi^2 \Lambda_S}{\rho_K} \right)^{n/2} \leq \frac{6 \log(2\pi^2) - 1}{6 \log(2\pi^2) - 2} \frac{e}{2\pi} < \frac{1}{2},$$

and (4) yields (1) with $\rho_{12} = 2\pi^2$. Moreover, if $S = \emptyset$, then $\Lambda_S = 1$ and

$$S_K \geq \gamma + \log(2\pi) - \frac{1}{6} - \frac{6 \log(2\pi^2) - 1}{6 \log(2\pi^2) - 2} \frac{\sqrt{e}}{\pi} \log(2\pi^2 \Lambda_S) > 0,$$

and (5) yields (2) with $\rho_{12} = 2\pi^2$.

(2) To deal with the cases $n \leq 5$, we use the following values: $J_1 = 2.62205 \dots$, $J_2 = \pi/2$, $J_3 = 1.19814 \dots$, $J_4 = 1$ and $J_5 = \frac{1}{3}J_1 = 0.87401 \dots$.

(3) For proving the last assertion of Theorem 1, we use (4) with $S = \{2\}$, and notice that $\Pi_K(\{2\}) \geq 1/(1 - 4^{-n})$, $(2\pi^2 \Lambda_S/\rho_K)^{n/2} \leq 4^{-n}$, $J_n \leq J_2 = \pi/2$ and

$$\frac{2eJ_n}{\pi^2} \frac{n \log \rho_K - 1}{n \log \rho_K - 2} \leq \frac{e}{\pi} \frac{2 \log(2700) - 1}{2 \log(2700) - 2} \leq 1.$$

3. LOWER BOUNDS FOR RELATIVE CLASS NUMBERS

Recall that a number field K is called a CM-field if K is totally imaginary, hence of even degree $2n \geq 2$, and if K is a quadratic extension of its maximal totally real subfield k . In that situation, the degree of k is equal to n , the class number h_k of k divides the class number h_K of K and we let $h_K^- = h_K/h_k$ denote the so-called relative class number of K . Recall that

$$(6) \quad h_K^- = \frac{Q_K w_K}{(2\pi)^n} \sqrt{\frac{d_K}{d_k}} \frac{\kappa_K}{\kappa_k}$$

where $w_K \geq 2$ and $Q_K \in \{1, 2\}$ are the number of complex roots of unity in K and the Hasse unit index of K , where d_K and d_k are the absolute values of the discriminants of the number fields K and k and where κ_K and κ_k are the residues at $s = 1$ of the Dedekind zeta functions of K and k (see [Was]). We finally let $\rho_K = d_K^{1/2n}$ and $\rho_k = d_k^{1/n}$ denote the root discriminants of K and k , respectively. Hence, $\rho_k \leq \rho_K$ and $d_K/d_k \geq d_K^{1/2} = \rho_K^n$.

Notation 5. Throughout this paper we adopt the following notation:

$$c_m = 2(\sqrt{m+1} - 1)^2.$$

$$(\text{In particular, } c_2 = 2(\sqrt{3} - 1)^2 = 1.07179 \dots \text{ and } c_3 = 2.)$$

$$\gamma = \text{Euler's constant} = 0.577215664901 \dots,$$

$$\kappa_1 = 2 + \gamma - \log(4\pi) = 0.046191417392 \dots,$$

$$\kappa_2 = 2 + \gamma - \log \pi = 1.432485779052 \dots,$$

$$\kappa_3 = 2 + \gamma - \log(\pi/4) = 2.818780140172 \dots.$$

$$\text{For } n \geq 2 \text{ we set } v_n = (n/(n-1))^{n-1} \in [2, e).$$

3.1. Upper bounds for residues of zeta functions. To obtain lower bounds for h_K^- , we will use (6), the lower bounds for κ_K obtained in Theorem 1 and the following upper bounds for κ_k :

Proposition 6.

(1) (See [Lou8, Theorem 1].) Let k be a number field of degree $n > 1$. Then

$$(7) \quad \kappa_k \leq \left(\frac{e \log d_k}{2(n-1)} \right)^{n-1} = v_n \left(\frac{e}{2} \log \rho_k \right)^{n-1}.$$

Moreover, if $\zeta_k(\beta) = 0$ for some β in the range $\frac{1}{2} \leq \beta < 1$, then

$$(8) \quad \kappa_k \leq (1-\beta) \left(\frac{e \log d_k}{2n} \right)^n = (1-\beta) \left(\frac{e}{2} \log \rho_k \right)^n.$$

(2) Let k be a real abelian number field of degree $n > 1$ and conductor $f_k > 1$. Let X_k denote the group (of order n) of primitive Dirichlet characters χ of conductors $f_\chi \geq 1$ associated with this abelian number field k (of degree n). Set

$$(9) \quad B_k := \prod_{1 \neq \chi \in X_k} \frac{1}{2} (\log f_\chi + \kappa_1) \leq \frac{v_n}{2^{n-1}} (\log \rho_k + \kappa_1)^{n-1}.$$

Then,

$$(10) \quad \kappa_k \leq B_k.$$

Moreover, if $\zeta_k(\beta) = 0$ for some β in the range $\frac{1}{2} \leq \beta < 1$, then

$$(11) \quad \kappa_k \leq \frac{(1-\beta) \log f_k}{4} B_k.$$

Proof. According to the conductor-discriminant formula, we do have

$$\begin{aligned} \prod_{1 \neq \chi \in X_k} (\log f_\chi + \kappa_1) &\leq \left(\frac{1}{n-1} \log d_k + \kappa_1 \right)^{n-1} \\ &\leq \left(\frac{n}{n-1} \log \rho_k + \frac{n}{n-1} \kappa_1 \right)^{n-1} = v_n (\log \rho_k + \kappa_1)^{n-1}. \end{aligned}$$

Now, $\kappa_k = \prod_{1 \neq \chi \in X_k} L(1, \chi) = \prod_{1 \neq \chi \in X_k} |L(1, \chi)|$. Hence, using Lemma 7 below, we obtain (10) and (11). \square

Lemma 7. Recall that we set $\kappa_1 = 2 + \gamma - \log(4\pi) = 0.046 \dots$. Let χ be a primitive even Dirichlet character of conductor $f_\chi > 1$.

(1) (See [Lou1].) We have¹

$$|L(1, \chi)| \leq (\log f_\chi + \kappa_1)/2.$$

(2) (See [Lou6, Corollary 7B] for the quadratic case, and [Lou8, Theorem 7] for the general case). Assume that $L(\beta, \chi) = 0$ for some $\beta \in [1/2, 1)$. Then,

$$|L(1, \chi)| \leq \frac{1-\beta}{8} \log^2 f_\chi.$$

3.2. On real zeros of Dedekind zeta functions.

Lemma 8. Set $c_m := 2(\sqrt{m+1} - 1)^2$ (hence, $\frac{1}{3} < c_1 < 1 < c_2 < c_3 = 2$).

(1) Let K be a normal CM-field. Set $c = c_3 = 2$ if K is abelian and $c = c_2 = 2(\sqrt{3} - 1)^2 = 1.07 \dots$ otherwise. Then, either (i) there exists some imaginary quadratic subfield F of K such that $\zeta_F(s)$ and $\zeta_K(s)$ have a common real zero in the range $1 - c/\log d_K \leq s < 1$, or (ii) $\zeta_k(s)$ has a real zero in the range $1 - c/\log d_K \leq s < 1$, or (iii) $\zeta_K(s) \leq 0$ in the range $1 - c/\log d_K \leq s < 1$.

(2) Let K be a not necessarily normal CM-field such that $(\zeta_K/\zeta_k)(s) \geq 0$ for $0 < s < 1$.² Then, either (i) $\zeta_k(s)$ has a real zero in the range $1 - 2/\log d_K \leq s < 1$, or (ii) $\zeta_K(s) \leq 0$ in the range $1 - 2/\log d_K \leq s < 1$.

(3) Let N denote the normal closure of a CM-field K . Then, either (i) there exists some imaginary quadratic subfield F of K such that $\zeta_F(s)$ and $\zeta_K(s)$ have a common real zero in the range $1 - (c_1/\log d_N) \leq s < 1$, or (ii) $\zeta_k(s)$ has a real zero in the range $1 - (c_1/\log d_N) \leq s < 1$, or (iii) $\zeta_K(s) \leq 0$ in the range $1 - (c_1/\log d_N) \leq s < 1$.

(4) Let N be the normal closure of a number field K of degree m . Then, d_N divides $d_K^{[N:\mathbf{Q}]}$ and the degree $[N:\mathbf{Q}]$ of N divides $m!$.

(5) Let F be an imaginary quadratic number field. Then, the Dedekind zeta function $\zeta_F(s)$ of F has no real zero in the range $1 - (6/\pi\sqrt{d_F}) \leq s < 1$.

(6) For any $\epsilon > 0$ there exists an ineffective constant $c_\epsilon > 0$ such that the Dedekind zeta functions $\zeta_F(s)$ of the imaginary quadratic number fields F have no real zero in the range $1 - (c_\epsilon/d_F^{\epsilon/2}) \leq s < 1$.

Proof. Let $m \geq 1$ be a given positive integer and K be a number field of discriminant $d_K > \exp(2(\sqrt{m+1} - 1))$. Then, its Dedekind zeta function $\zeta_K(s)$ has at most m real zeros in the range $1 - (c_m/\log d_K) \leq s < 1$ (this result is a generalisation of [Sta3, Lemma 3] and its proof is given in [LLO, Lemma 15]).

¹We could choose $\kappa_1 = 0$, by [Ram, Corollary 1].

²For example, K is a dihedral or quaternion CM-field of degree $2n \geq 8$.

(1) The abelian case is easy to deal with by using the factorization of $\zeta_K(s)$ as a product of Dirichlet L -series. Let us now deal with the normal case. Assume that we are neither in case (i) nor in case (iii). Since we are not in case (iii), there exists s_1 in the range $1 - c_2/\log d_K \leq s < 1$ such that $\zeta_K(s_1) > 0$. Since κ_K is positive and since $\zeta_K(s)$ has a simple pole at $s = 1$, it follows that $\lim_{s \uparrow 1} \zeta_K(s) = -\infty$. Hence, $\zeta_K(s)$ has a real zero β of odd multiplicity $n_\beta \geq 1$ in the range $1 - c_2/\log d_K \leq s_1 \leq s < 1$. However, in this range we have $n_\beta \leq 2$. Hence, $n_\beta = 1$. According to [Sta3, Theorem 3], there exists some quadratic subfield F of K such that for $E \subseteq K$ we have $\zeta_E(\beta) = 0$ if and only if $F \subseteq E$. In particular, $\zeta_F(\beta) = 0$. Since we are not in case (i), we obtain that F is real. Hence, $F \subseteq k$, which implies $\zeta_k(\beta) = 0$, and we are in case (ii).

(2) Easy.

(3) Assume that we are neither in case (i) nor in case (iii). Since we are not in case (iii), there exists s_1 in the range $1 - (c_1/\log d_K) \leq s < 1$ such that $\zeta_K(s_1) > 0$. Since $\lim_{s \uparrow 1} \zeta_K(s) = -\infty$, there exists some real zero β of ζ_K in the range $s_1 \leq s < 1$. Since N/K is normal, ζ_K divides ζ_N (see [FM, Corollary 2] for a short proof of the Aramata-Brauer Theorem). Hence, $\zeta_N(\beta) = 0$ and β is a simple zero of ζ_N , for $1 - (c_1/\log d_N) \leq s_1 \leq \beta < 1$. According to [Sta3, Theorem 3], there exists a quadratic subfield $F \subseteq N$ such that $E \subseteq N$ and $\zeta_E(\beta) = 0$ if and only if $F \subseteq E$. In particular, $F \subseteq K$ (for $\zeta_K(\beta) = 0$) and $\zeta_F(\beta) = 0$. Since we are not in case (i), then F is real. Hence $F \subseteq k$ and $\zeta_k(\beta) = 0$ and we are in case (ii).

(4) Use [Sta3, Lemma 7].

(5) See [Bes] (the proof of this result was announced to appear in [Hof, Reference 2] but it has in fact never been published yet).

(6) See [Pin1, Siegel’s Theorem II], [Pin2, Theorem 1] and [Sie]. □

3.3. Lower bounds for relative class numbers. We are now in a position to obtain lower bounds for relative class numbers.

Theorem 9. *Let K be a normal CM-field of degree $2n > 2$ and root discriminant $\rho_K \geq 50$. Assume that K contains no imaginary quadratic subfield, or that the Dedekind zeta functions of the imaginary quadratic subfields of K have no real zero in the range $1 - (c/\log d_K) \leq s < 1$.*

(1) Set $c = c_2 = 2(\sqrt{3} - 1)^2$. We have

(12)

$$h_K^- \geq \frac{cQ_Kw_K\sqrt{d_K/d_k}}{2\pi e^{c/2}v_n(\pi e \log \rho_k)^{n-1} \log d_K} \geq \frac{c}{2nv_n e^{c/2-1}} \left(\frac{\sqrt{\rho_K}}{\pi e \log \rho_K} \right)^n,$$

and for each entry $2m$ in Table 2 below, $2n \geq 2m$ and $\rho_K \geq \rho_{2m}$ imply $h_K^- > 1$.

(2) Moreover, assume that k is abelian and set $c = c_3 = 2$ if K is abelian, and set $c = c_2 = 2(\sqrt{3} - 1)^2$ otherwise. Then, we have the better lower bound

(13)

$$h_K^- \geq \frac{cQ_Kw_K\sqrt{d_K/d_k}}{e^{c/2}(2\pi)^n B_k \log d_K} \geq \frac{c}{2nv_n e^{c/2}} \left(\frac{\sqrt{\rho_K}}{\pi(\log \rho_K + \kappa_1)} \right)^n$$

(recall that we have set $\kappa_1 = 2 + \gamma - \log(4\pi) = 0.046 \dots$), and for each entry $2m$ in Table 3 below, $2n \geq 2m$ and $\rho_K \geq \rho_{2m}$ imply $h_K^- > 1$.

Table 2 : $c = c_2$

$2m$	4	6	8	10	20	40	100	200	∞
ρ_{2m}	38100	31000	25000	21000	13000	9200	7000	6260	5383

Table 2 (continued) : $c = 2$

$2m$	4	6	8	10	20	40	100	200	∞
ρ_{2m}	31300	27200	22600	19400	12500	9000	7000	6230	5383

Table 3 : k is abelian and $c = c_2$

$2m$	4	6	8	10	20	40	100	200	∞
ρ_{2m}	11100	5800	3710	2700	1220	726	490	418	342

Table 3 (continued) : k is abelian and $c = 2$

$2m$	4	6	8	10	20	40	100	200	∞
ρ_{2m}	9000	5100	3340	2480	1170	709	486	416	342

Proof. Let us first prove (13). According to Point 1 of Lemma 8, there are two cases to consider.

(1) Assume that ζ_k has no real zero in the range $1 - c/\log d_K \leq s < 1$. Then $\zeta_K(1 - (c/\log d_K)) \leq 0$ and using (2) with $S = \emptyset$, we obtain

$$\kappa_K \geq \frac{c}{e^{c/2} \log d_K}.$$

Using (10) we conclude that

$$(14) \quad \frac{\kappa_K}{\kappa_k} \geq \frac{c}{e^{c/2} B_k \log d_K}.$$

(2) Assume that $\zeta_k(\beta) = 0$ for some $\beta \in [1 - c/\log d_K, 1)$. Then $\zeta_K(\beta) = 0 \leq 0$ and using (1) with $S = \emptyset$, we obtain

$$\kappa_K \geq \frac{1 - \beta}{2e^{c/2}}.$$

Using (11) we conclude that

$$(15) \quad \frac{\kappa_K}{\kappa_k} \geq \frac{2}{e^{c/2} B_k \log f_k}.$$

Since $d_K > d_k \geq f_k$ and since $c \leq 2$, the right-hand side of (15) is greater than or equal to the right-hand side of (14), and (14) is always valid.

Using (14), (9) and (6), we obtain the first lower bound in (13). To deduce the second one, we use $\log d_K = 2n \log \rho_K$, $\sqrt{d_K/d_k} \geq d_K^{1/4} = \rho_K^{n/2}$ and $\rho_k \leq \rho_K$.

To prove (12), we use (7) and (8), instead of (10) and (11). \square

Remarks 10. According to (4), we could easily improve upon (1). For example, we have: let K be a totally imaginary number field of degree $2n \geq 4$ and root discriminant $\rho_K := d_K^{1/2n} \geq 98$. Assume that $\zeta_K(\beta) \leq 0$ for some $\beta \in [1 - (2/\log d_K), 1)$. Then,

$$(16) \quad \kappa_K \geq \frac{4}{5}(1 - \beta)e^{(\beta-1)/2}.$$

The reader can easily check that by following the proof of Theorem 9 and by using (3), or (16) with $S = \emptyset$, we can slightly improve upon [LPP, Proposition 4.2].

In the same way, by using Point 2 of Lemma 8 we obtain:

Theorem 11. *Let K be a not necessarily normal CM-field of degree $2n \geq 2$ such that $(\zeta_K/\zeta_k)(s) \geq 0$ for $0 < s < 1$.³ Then, (12) holds with $c = 2$. In particular, for each entry $2m$ in Table 2, we have $h_K^- > 1$ as soon as $2n \geq 2m$ and $\rho_K \geq \rho_{2m}$. Moreover, if k is abelian,⁴ then (13) holds with $c = 2$. In particular, for each entry $2m$ in Table 3, we have $h_K^- > 1$ as soon as $n \geq m$ and $\rho_K \geq \rho_{2m}$.*

Finally, by using (3), (7), (8) and Points 3 and 4 of Lemma 8 we obtain:

Theorem 12. *Set $c = c_1 = 2(\sqrt{2}-1)^2$. Let K be a not necessarily normal CM-field of degree $2n > 2$ and root discriminant $\rho_K \geq 2800$, let $m_N := [N : \mathbf{Q}]$ denote the degree of its normal closure N and assume that K contains no imaginary quadratic subfield or that the real zeros in the range $1 - (c/\log d_N) \leq s < 1$ of the Dedekind zeta functions of the imaginary quadratic subfields of K are not zeros of $\zeta_K(s)$. Then,*

(17)
$$h_K^- \geq \frac{cQ_K w_K \sqrt{d_K/d_k}}{4nm_N e^{c/2} (\pi e \log \rho_K)^n}.$$

When dealing with small class number problems for CM-fields K , one can assume that either K contains no imaginary quadratic subfield or that $\zeta_F(s) < 0$ in the range $0 < s < 1$ for all the imaginary quadratic subfields F of K , which enables one to use Theorems 9 and 12. Indeed, the class number of any imaginary quadratic subfield of K divides $4h_K^-$ (see [Oka]), all the imaginary quadratic fields of small class numbers are known (e.g. those of class numbers dividing 4 were determined in [Arn], [Bak1], [Bak2], [MW], [Sta1] and [Sta2]), and one can easily check numerically that $\zeta_F(s) < 0$, $0 < s < 1$, for these few imaginary quadratic fields F . However, in order to prove in Section 4 a Brauer-Siegel-like result for relative class numbers of CM-fields, we prove:

Theorem 13. *Let \gg_ϵ mean that the constants involved in the considered lower bound depend on ϵ only. Let K be a not necessarily normal CM-field of degree $2n > 4$ and root number $\rho_K \geq 2800$. Assume that K contains an imaginary quadratic subfield F and that $\zeta_F(\beta) = \zeta_K(\beta) = 0$ for some $\beta \in [-(2/\log d_K), 1)$. Then,*

(18)
$$h_K^- \geq \frac{6}{\pi^2 e^2} \frac{(d_K/d_k)^{\frac{1}{2}-\frac{1}{n}}}{(\pi e \log \rho_K)^{n-1}} \geq \frac{6}{\pi^2 e^2 \sqrt{\rho_K}} \left(\frac{\sqrt{\rho_K}}{\pi e \log \rho_K} \right)^{n-1}$$

and (ineffectively)

(19)
$$h_K^- \gg_\epsilon \frac{(d_K/d_k)^{\frac{1}{2}-\frac{\epsilon}{n}}}{(\pi e \log \rho_K)^{n-1}},$$

and for each entry $2m$ in Table 4 below, $2n \geq 2m$ and $\rho_K \geq \rho_{2m}$ imply $h_K^- > 1$.

Table 4: $c = c_2$

$2m$	6	8	10	12	20	40	100	200	∞
ρ_{2m}	$5 \cdot 10^{11}$	$5 \cdot 10^7$	$3 \cdot 10^6$	$5 \cdot 10^5$	50000	15000	7800	6500	5383

According to (12) and (18), it follows that $h_K^- \rightarrow \infty$ as $[K : \mathbf{Q}] = 2n \rightarrow \infty$ for normal CM-fields of root discriminants $\rho_K = d_K^{1/2n} \geq c_\infty := 5400$.

³For example, K is a dihedral CM-field of degree $2n \geq 8$ as in [LO].
⁴For example, K is a quaternion or a dihedral octic CM-field as in [Lou3] and [Lou5, Section 2.3.1].

Proof. According to Point 4 of Lemma 8, we have $1 - \beta \geq \frac{6}{\pi}(d_K/d_k)^{-1/n}$ and $1 - \beta \gg_\epsilon (d_K/d_k)^{-\epsilon/n}$ (notice that $\sqrt{d_F} = \rho_F \leq \rho_K \leq (d_K/d_k)^{1/n}$). According to (3), we have $\kappa_K \geq (1 - \beta)/e$, and we obtain

$$(20) \quad \kappa_K \geq \frac{6}{\pi e}(d_K/d_k)^{-1/n} \quad \text{and} \quad \kappa_K \gg_\epsilon (d_K/d_k)^{-\epsilon/n}.$$

Using (6), (7) and (20), we obtain (18) and (19). \square

4. A BRAUER-SIEGEL-LIKE RESULT ON THE ASYMPTOTIC BEHAVIOR OF RELATIVE CLASS NUMBERS OF CM-FIELDS

By using our previous lower bounds for relative class numbers of CM-fields (see Theorems 9, 12 and 13), we now prove Brauer-Siegel-like results about the asymptotic behavior of relative class numbers of CM-fields. In [HJ, p. 554] it is said that the restriction $\rho_K \rightarrow \infty$ precludes one from deducing from the Brauer-Siegel theorem that there exists some sufficiently large constant $C > 0$ such that $h_K \rightarrow \infty$ as $[K : \mathbf{Q}] = 2n \rightarrow \infty$ for normal CM-fields K of root discriminants $\rho_K = d_K^{1/2n} \geq C$. The Brauer-Siegel-like results we will obtain here prove that we may choose $C = 5400$. In [Mur2, Proposition 4.1] it is said that as K ranges over the set of CM-fields of degrees $2n \leq 8$ and $2n \neq 4$ we have $h_K \rightarrow \infty$ effectively. The Brauer-Siegel-like results we will obtain here prove that for any given B we have $h_K \rightarrow \infty$ effectively as K ranges over the set of CM-fields of degrees $2n \leq B$. The Brauer-Siegel-like results for relative class number of CM-fields we are going to prove (and which generalize those we obtained in [Lou4] for imaginary abelian number fields) are as follows:

Theorem 14.

(1) *Let K range over a sequence of normal CM-fields such that their root discriminants ρ_K tend to infinity (e.g. let K range over a sequence of imaginary abelian number fields⁵), or let K range over a sequence of not necessarily normal CM-fields of a given degree. Let $o(1)$ denote an error term that tends to zero as ρ_K goes to infinity.*

We have

$$(21) \quad \left(\frac{1}{2} + o(1)\right) \log(d_K/d_k) \geq \log h_K^- \geq \left(\frac{1}{2} + o(1)\right) \log(d_K/d_k),$$

i.e., $\log h_K^-$ is asymptotic to $\frac{1}{2} \log(d_K/d_k)$, which implies that

$$(22) \quad h_K^- \gg d_K^{\frac{1}{4} + o(1)}.$$

The upper bound on $\log h_K^-$ in (21) is effective and explicit.

If K contains no imaginary quadratic subfield, then the lower bounds for $\log h_K^-$ in (21) and (22) are effective and explicit.

If K contains an imaginary quadratic subfield, then the lower bounds for h_K^- in (21) and (22) are not effective, but we have the following effective and explicit weaker lower bound:

$$(23) \quad \log h_K^- \geq \left(\frac{1}{2} - \frac{1}{n} + o(1)\right) \log(d_K/d_k),$$

⁵For in that case it follows that $\rho_K \geq \sqrt{f_K}$ (see [Mur1, Corollary 1]).

which implies the following effective and explicit lower bound:

$$(24) \quad h_K^- \gg d_K^{\frac{1}{4} - \frac{1}{2n} + o(1)}.$$

Finally, in the situations where the error terms $o(1)$ in (21), (22), (23) and (24) are declared to be effective and explicit, they are of the type $o(1) = O((\log \log \rho_K) / \log \rho_K)$.

(2) If K ranges over not necessarily normal CM-fields of a given degree, then $h_K^- \rightarrow \infty$ effectively and explicitly as $d_K \rightarrow \infty$.

For any given $h \geq 1$ there exists ρ_h effective such that $h_K^- > h$ for all normal CM-fields K of root discriminants $\rho_K \geq \rho_h$.

In particular, $h_K^- > 1$ for all normal CM-fields K of root discriminants $\rho_K \geq \rho_1 = 40000$.

Moreover, $h_K \rightarrow \infty$ as $[K : \mathbf{Q}] = 2n \rightarrow \infty$ for normal CM-fields K of root discriminants $\rho_K = d_K^{1/2n} \geq C = 5400$.

4.1. Proof of Theorem 14.

Lemma 15. *Let K be a CM-field of degree $2n$. Then,*

$$(25) \quad \log h_K^- \leq \left(\frac{1}{2} + o(1) \right) \log(d_K/d_k)$$

where $o(1) = O((\log \log \rho_K) / \log \rho_K)$ is an explicit error term that tends to zero as ρ_K goes to infinity.

Proof. Since $\phi(w) \geq \sqrt{w/2}$ for $w \geq 2$ and since $\phi(w_K)$ must divide $2n$, we have $w_K \leq 8n^2$. Moreover, $d_K/d_k \leq d_K = \rho_K^{2n}$. Hence, using [Lou7, Corollary 3], we obtain

$$h_K^- \leq 2Q_K w_K \sqrt{d_K/d_k} \left(\frac{e}{4\pi n} \log(d_K/d_k) \right)^n \leq 32n^2 \sqrt{d_K/d_k} \left(\frac{e}{2\pi} \log \rho_K \right)^n$$

and the desired result, by using $\log(d_K/d_k) \geq \log(d_K^{1/4}) = n \log \rho_K$. \square

1. The first point of Theorem 14 follows from Lemma 15 and Theorems 9, 12 and 13 (to prove the last assertion of the first point of Theorem 14, recall that $\log(d_K/d_k) \geq \log(d_K^{1/2}) = n \log \rho_K$).

2. The first and second assertions of the second point of Theorem 14 follow from the first point of Theorem 14 (for CM-fields of degrees $2n > 4$) and from the following known results (for CM-fields of degrees $2n \leq 4$):

Lemma 16.

(1) (See [Oes].) For every $\epsilon > 0$ we have an effective and explicit lower bound $h_F^- \gg_\epsilon \log^{1-\epsilon} d_F$ for the class numbers h_F of the imaginary quadratic fields F .

(2) If $K = F_1 F_2$ is an imaginary bicyclic biquadratic field (where F_1 and F_2 denote the two imaginary quadratic subfields of K), then $d_K/d_k = d_{F_1} d_{F_2}$ and

$$h_K^- = \frac{Q_K}{2} h_{F_1} h_{F_2}.$$

Hence, we have an effective and explicit lower bound $h_K^- \gg_\epsilon \log^{1-\epsilon} d_K$.

(3) If K is a non-normal quartic CM-field, then its normal closure N is a dihedral octic CM-field, $d_N/d_{N^+} = (d_K/d_k)^2$, and

$$h_N^- = \frac{Q_N}{2} (h_K^-)^2.$$

Therefore, $\log h_K^-$ is effectively and explicitly asymptotic to $\frac{1}{2} \log(d_K/d_k)$.

3. The third assertion of the second point of Theorem 14 follows from the fact that if $h_K^- = 1$ and K contains an imaginary quadratic field F , then h_F divides 4 (see [Oka]). Hence F is known (see [Arn], [Bak1], [Bak2], [MW], [Sta1] and [Sta2]), and numerical computations easily yield that $\zeta_F(s) < 0$ for these few imaginary quadratic fields F . Hence, the first point of Theorem 9 yields that $h_K^- > 1$ for all normal CM-fields K of root discriminants $\rho_K > 40000$. (We could also use Theorem 13 and the solution of the class number one problem for the imaginary quadratic fields (see [Bak1] and [Sta1]) and for the imaginary biquadratic bicyclic fields (see [BP]), but we would obtain the weaker following result: $h_K^- > 1$ for all normal CM-fields K of root discriminants $\rho_K > 7 \cdot 10^{11}$.)

4. Finally, the fourth assertion of the second point of Theorem 14 follows from the last assertion of Theorem 13.

Remarks 17. It is possible to deduce from the usual Brauer-Siegel theorem for class numbers of number fields the following Brauer-Siegel-like result for relative class numbers of normal CM-fields, which improves upon [HH, Lemma 4] (which is given only for CM-fields of a given degree) but is less satisfactory than our previous Theorem 14 (for it is ineffective in the case that N contains no imaginary quadratic subfield):

Theorem 18. *If N ranges over a sequence of normal CM-fields such that their root discriminants ρ_N tend to infinity, then we have*

$$\log h_N^- \sim \frac{1}{2} \log(d_N/d_{N+}),$$

which implies

$$h_N^- \gg d_N^{\frac{1}{4} + o(1)}$$

where $o(1)$ is an error term that tends to zero as ρ_N goes to infinity.

5. BETTER LOWER BOUNDS FOR RELATIVE CLASS NUMBERS

The aim of this section is to improve upon, in the case that k is abelian, the explicit lower bounds for relative class numbers of CM-fields K that we obtained in the previous section. To this end, we choose $S = \{2\}$ and use Theorem 1 to get better lower bounds (depending on the behavior of 2 in K) for the term κ_K in (6). Moreover, using the results of [Lou9] we will be able to get better upper bounds (depending on the behavior of 2 in k) for the term κ_k in (6). Putting everything together, we will obtain Theorem 22, which improves upon the lower bounds for relative class numbers that we obtained in Theorem 9.

5.1. Upper bounds for $|L(1, \chi)|$.

Lemma 19. *Let χ be a primitive even Dirichlet character of conductor $f_\chi > 1$.*

(1) (See [Lou1] and [Lou9].) Set ⁶

$$(26) \quad \kappa_\chi := \begin{cases} \kappa_1 = 2 + \gamma - \log(4\pi) = 0.046 \dots & \text{if } \chi(2) = +1, \\ \kappa_2 = 2 + \gamma - \log \pi = 1.432 \dots & \text{if } \chi(2) = 0, \\ \kappa_3 = 2 + \gamma - \log(\pi/4) = 2.818 \dots & \text{if } \chi(2) \neq 0, +1. \end{cases}$$

⁶We could choose $\kappa_1 = 0$ and $\kappa_2 = \log 4 = 1.386 \dots$, by [Ram, Corollaries 1 and 2].

We have

$$(27) \quad |L(1, \chi)| \leq \frac{1}{4} \left| 1 - \frac{\chi(2)}{2} \right|^{-1} (\log f_\chi + \kappa_\chi).$$

(2) (See [Lou6, Corollary 7B] for the quadratic case and [Lou8, Theorem 7] for the general case.) If $L(\beta, \chi) = 0$ for some $\beta \in [1/2, 1)$, then

$$(28) \quad |L(1, \chi)| \leq \frac{1-\beta}{8} \log^2 f_\chi.$$

5.2. Upper bounds for residues of zeta functions.

Proposition 20. Let k be a real abelian number field of degree $n > 1$ and conductor $f_k > 1$. Let X_k denote the group (of order n) of primitive Dirichlet characters χ of conductors $f_\chi \geq 1$ associated with this abelian number field k (of degree n). Let e , f and $g = n/(ef)$ denote the index of ramification of 2 in k , the inertia degree of 2 in k and the number of prime ideals of k above 2, respectively. (Hence, $\Pi_k(\{2\}) = (1 - 2^{-f})^{-g}$.) Set

$$(29) \quad \kappa_k := \frac{1}{n} \sum_{1 \neq \chi \in X_k} \kappa_\chi \quad (\text{with } \kappa_\chi \text{ as in (26)})$$

$$(30) \quad \leq \kappa_{n,f,g} := \frac{(g-1)\kappa_1 + (n-fg)\kappa_2 + (fg-g)\kappa_3}{n}$$

(hence $0 < \kappa_k \leq \kappa_{n,f,g} \leq \kappa_3 \leq 3$) and

$$(31) \quad B_k(\{2\}) := \frac{1}{2} \prod_{1 \neq \chi \in X_k} \frac{1}{4} (\log f_\chi + \kappa_\chi) \leq \frac{v_n}{2 \cdot 4^{n-1}} (\log \rho_k + \kappa_k)^{n-1}.$$

Then,

$$(32) \quad \kappa_k \leq \Pi_k(\{2\}) B_k(\{2\}).$$

Moreover, if $\zeta_k(\beta) = 0$ for some $\beta \in [1/2, 1)$, then

$$(33) \quad \kappa_k \leq \frac{3(1-\beta) \log f_k}{4} \Pi_k(\{2\}) B_k(\{2\}).$$

Proof. To deduce (30) from (29), we notice that, according to [Was, Theorem 3.7], we have $\#\{\chi \in X_k; \chi(2) = 1\} = g$, $\#\{\chi \in X_k; \chi(2) = 0\} = n - fg$ and $\#\{\chi \in X_k; \chi(2) \neq 0, 1\} = n - g - (n - fg) = fg - g$. Using the fact that the geometric mean is less than or equal to the arithmetic mean and the conductor-discriminant formula $\prod_{1 \neq \chi \in X_k} f_\chi = d_k = \rho_k^n$, we do have

$$B_k(\{2\}) \leq \frac{1}{2 \cdot 4^{n-1}} \left(\frac{1}{n-1} \sum_{1 \neq \chi \in X_k} (\log f_\chi + \kappa_\chi) \right)^{n-1} = \frac{v_n}{2 \cdot 4^{n-1}} (\log \rho_k + \kappa_k)^{n-1}.$$

Noticing that

$$\prod_{1 \neq \chi \in X_k} \left(1 - \frac{\chi(2)}{2} \right)^{-1} = \frac{1}{2} \Pi_k(\{2\})$$

and using (27) for all the $1 \neq \chi \in X_k$, we obtain

$$\kappa_k = \prod_{1 \neq \chi \in X_k} |L(1, \chi)| \leq \frac{\Pi_k(\{2\})}{2 \cdot 4^{n-1}} \prod_{1 \neq \chi \in X_k} (\log f_\chi + \kappa_\chi),$$

which proves (32). Now, if $\zeta_k(\beta) = 0$, then $L(\beta, \chi_0) = 0$ for some $1 \neq \chi_0 \in X_k$. Using (28) we obtain

$$\begin{aligned} |L(1, \chi_0)| &\leq \frac{1-\beta}{8} \log^2 f_{\chi_0} \leq \frac{1-\beta}{8} \cdot \frac{3}{2} \left|1 - \frac{\chi_0(2)}{2}\right|^{-1} \cdot \log^2 f_{\chi_0} \\ &\leq \frac{3(1-\beta) \log f_k}{4} \left|1 - \frac{\chi_0(2)}{2}\right|^{-1} \frac{\log f_{\chi_0} + \kappa_{\chi_0}}{4}, \end{aligned}$$

which, in using (27) for all the $\chi \in X_k \setminus \{1, \chi_0\}$, yields (33). □

Remarks 21. Notice that in the special case that the prime 2 is inert in the real abelian number field k of degree n , then (31) and (32) yield

$$\kappa_k \leq \frac{v_n}{2^{n-1}(2^n - 1)} (\log \rho_k + \kappa_3)^{n-1},$$

whereas (9) and (10) only yield $\kappa_k \leq v_n (\log \rho_k + \kappa_1)^{n-1} / 2^{n-1}$.

5.3. Lower bounds for relative class numbers.

Theorem 22. *Let K be a normal CM-field of degree $2n \geq 2m > 2$ and root discriminant $\rho_K \geq \rho_{2m, \{2\}}$ with $\rho_{2m, \{2\}}$ as in Table 1. Assume that k is abelian. Set $c = c_3 = 2$ if K is abelian, and set $c = c_2 = 2(\sqrt{3} - 1)^2$ otherwise. Assume that K contains no imaginary quadratic subfield or that the Dedekind zeta functions of the imaginary quadratic subfields of K have no real zero in the range $1 - (c/\log d_K) \leq s < 1$. Then,*

$$(34) \quad h_K^- \geq \frac{c}{e^{c/2}} \frac{Q_K w_K \Pi_{K/k}(\{2\}) \sqrt{d_K/d_k}}{(2\pi)^n B_k(\{2\}) \log d_K}$$

with $B_k(\{2\})$ as in (31) and $\Pi_{K/k}(\{2\}) = \Pi_K(\{2\})/\Pi_k(\{2\})$.

Therefore, setting $C_{n,f,g} = 2(1 + 2^{-f})^{-g/n} \in [4/3, 2)$, we have

$$(35) \quad h_K^- \geq \frac{c}{2nv_n e^{c/2}} \left(\frac{C_{n,f,g} \sqrt{\rho_K}}{\pi(\log \rho_K + \kappa_{n,f,g})} \right)^n$$

(with f, g and $\kappa_{n,f,g}$ as in Proposition 20). In particular, for each entry $2n$ in Table 6 below, we have $h_K^- > 1$ as soon as $\rho_K \geq \rho_{2n}$.

Table 6 (compare with Table 3)

$2n$	4	6	8	10	20	40	100	200
ρ_{2n} for $c = c_2$	5217	2704	1707	1228	538	310	206	181
ρ_{2n} for $c = 2$	4233	2344	1530	1124	513	303	205	180

Proof. According to Points 1 and 2 of Lemma 8, there are two cases to consider.

(1) Assume that ζ_k has no real zero in the range $1 - c/\log d_K \leq s < 1$. Then $\zeta_K(1 - (c/\log d_K)) \leq 0$ and using (2) with $S = \{2\}$, we obtain

$$\kappa_K \geq \frac{c \Pi_K(\{2\})}{e^{c/2} \log d_K}.$$

Using (32) we conclude that

$$(36) \quad \frac{\kappa_K}{\kappa_k} \geq \frac{c \Pi_{K/k}(\{2\})}{e^{c/2} B_k(\{2\}) \log d_K}.$$

(2) Assume that $\zeta_k(\beta) = 0$ for some $\beta \in [1 - (c/\log d_K), 1)$. Then $\zeta_K(\beta) = 0 \leq 0$ and using (1) with $S = \{2\}$, we obtain

$$\kappa_K \geq \frac{(1 - \beta)\Pi_{K/k}(\{2\})}{2e^{c/2}}.$$

Using (33) we conclude that

$$(37) \quad \frac{\kappa_K}{\kappa_k} \geq \frac{c\Pi_{K/k}(\{2\})}{e^{c/2}B_k(\{2\})^{\frac{3c}{2}}\log f_k}.$$

Now, if $n \geq 3$, then $d_K \geq d_k^2 \geq f_k^3$ (see [Mur1, Corollary 1]), and if $n = 2$, then K is cyclic quartic and here again $d_K = f_K^2 f_k \geq f_k^3$. Hence, we always have

$$\frac{3c}{2} \log f_k \leq 3 \log f_k \leq \log d_K$$

(for $c \leq 2$). Therefore, the right-hand side of (37) is greater than or equal to the right-hand side of (36), and (36) is always valid. Using (36), (31) and (6), we obtain (34). To deduce (35), we use $\log d_K = 2n \log \rho_K$, $\sqrt{d_K/d_k} \geq d_K^{1/4} = \rho_K^{n/2}$, $\rho_k \leq \rho_K$ and

$$(38) \quad 2^n \Pi_{K/k}(\{2\}) = 2^n (1 - \epsilon_2/2^f)^{-g} \geq (C_{n,f,g})^n$$

(where $\epsilon_2 = -1, 0$ or 1 according as the prime ideals of k above 2 are inert, ramified or split in the quadratic extension K/k). In particular, $\Pi_{K/k}(\{2\}) = 1$ if 2 is ramified in K/k . Finally, since $\kappa_{n,f,g} > 0$, the right-hand side of (35) increases with $\rho_K \geq e^2$. Hence, for a given n and a given $\rho_K \geq 55 > e^4$ we can easily compute the minima of the right-hand sides of (35) over all the pairs (f, g) with $f \geq 1$ and $n \geq 1$ such that fg divides n , and these minima increase with $\rho_K \geq 55 > e^4$. This makes it easy to compute ρ_{2n} for any given entry $2n$ in Table 6. \square

In the same way, by using Point 2 of Lemma 8 we also obtain:

Theorem 23. *Let K be a not necessarily normal CM-field of degree $2n \geq 2$ such that $(\zeta_K/\zeta_k)(s) \geq 0$ for $0 < s < 1$ and such that k is abelian.⁷ Then (34) and (35) hold with $c = 2$. In particular, for each entry $2n$ in Table 6, we have $h_K^- > 1$ as soon as $\rho_K \geq \rho_{2n}$.*

6. AN APPLICATION OF THESE BETTER LOWER BOUNDS

The aim of this section is to give an example showing the paramount usefulness of Theorem 22 when dealing with class group problems for various types of CM-fields for which Theorem 9 is of less or no practical usefulness. In [Lou5] we proved that if K is a non-normal quartic CM-field, then

$$(39) \quad h_K^- \geq \frac{\sqrt{d_K/d_k}}{12(\log(d_K/d_k) + 0.052)^2}$$

(notice that according to its proof, there is a misprint in the statement of the lower bound [Lou5, Corollary 15]). We will now improve upon this lower bound.

⁷For example, K is a quaternion or a dihedral octic CM-field as in [Lou3], and [Lou5, Section 2.3.1].

Lemma 24. *Let N be the normal closure of a non-normal quartic CM-field K . Hence, N is a dihedral octic CM-field. Then,*

$$(40) \quad h_N^- \geq \frac{Q_N \Pi_{N/N^+}(\{2\}) \sqrt{d_N/d_{N^+}}}{4e\pi^4 B_{N^+}(\{2\}) \log d_N}$$

(for $\rho_N \geq 222$), and

$$(41) \quad B_{N^+}(\{2\}) \log d_N \leq (\log(d_K/d_k) + 3)^4/128.$$

Proof. To get (40), use (34) with $c = 2$ (see Theorem 23). Let us now prove (41). Let $L_1 = k$, L_2 and L_+ be the three real quadratic subfields of N^+ , the extension N/L_+ being cyclic quartic, and let λ_1 , λ_2 and λ_+ be the constants κ_χ defined in (26) associated with the three quadratic characters χ of these three real quadratic fields. It is known that $L_2 = \mathbf{Q}(\sqrt{d_K/d_k^2})$ and that d_{L_2} divides d_K/d_k^2 . Since d_{L_+} divides the product $d_{L_1}d_{L_2}$ (for N^+/\mathbf{Q} is biquadratic bicyclic), we conclude that d_{L_+} divides d_K/d_k . Upon using the bound $d_N \leq (d_N/d_{N^+})^2 = (d_K/d_k)^4$, we obtain (see (31)):

$$\begin{aligned} 128B_{N^+}(\{2\}) \log d_N &\leq 4(\log d_k + \lambda_1)(\log(d_K/d_k^2) + \lambda_2)(\log(d_K/d_k) + \lambda_+) \log(d_K/d_k) \\ &\leq (\log(d_K/d_k) + \lambda_1 + \lambda_2)^2 (\log(d_K/d_k) + \lambda_+) \log(d_K/d_k) \\ &\quad (\text{for } 4ab \leq (a+b)^2 \text{ for } a \geq 0 \text{ and } b \geq 0) \\ &\leq (\log(d_K/d_k) + (2\lambda_1 + 2\lambda_2 + \lambda_+)/4)^4 \\ &\quad (\text{for } a^2bc \leq ((2a+b+c)/4)^4 \text{ for } a \geq 0, b \geq 0 \text{ and } c \geq 0). \end{aligned}$$

Finally, since either 2 splits in one of the three quadratic subfields of k , or 2 ramifies in at least two of the three quadratic subfields of k , we have $(2\lambda_1 + 2\lambda_2 + \lambda_+)/4 \leq (4\kappa_3 + \kappa_1)/4 = 2.830327 \dots$. \square

Theorem 25. *Let K be a non-normal quartic CM-field. Assume that $\rho_K \geq 222$. Then,*

$$(42) \quad h_K^- \geq \frac{8\Pi_{K/k}(\{2\})\sqrt{d_K/d_k}}{\sqrt{e}\pi^2(\log(d_K/d_k) + 3)^2} \geq \frac{\sqrt{d_K/d_k}}{C_K(\log(d_K/d_k) + 3)^2}$$

where

$$C_K = \begin{cases} 9\sqrt{e}\pi^2/32 = 4.57656\dots & \text{if 2 is not ramified in } K, \\ 3\sqrt{e}\pi^2/16 = 3.05104\dots & \text{if 2 is ramified in } K, \\ \sqrt{e}\pi^2/8 = 2.03402\dots & \text{if 2 is totally ramified in } K. \end{cases}$$

Proof. Let N denote the normal closure of K . Then N is a dihedral octic CM-field. Since $\zeta_N/\zeta_{N^+} = (\zeta_K/\zeta_k)^2$, it follows that $d_N/d_{N^+} = (d_K/d_k)^2$, $\Pi_{N/N^+}(\{2\}) = (\Pi_{K/k}(\{2\}))^2$ and $h_N^- = Q_N(h_K^-)^2/2$. Using (40) and Lemma 24, we obtain the first lower bound for h_K^- .

As for the second lower bound, we use

$$\Pi_{K/k}(\{2\}) = \prod_{\mathcal{P}_k | (2)} \left(1 - \frac{\chi(\mathcal{P}_k)}{N_{k/\mathbf{Q}}(\mathcal{P}_k)}\right)^{-1} \geq \begin{cases} 1 & \text{if 2 is totally ramified in } K, \\ 2/3 & \text{if 2 is ramified in } K, \\ (2/3)^2 & \text{in all cases,} \end{cases}$$

where \mathcal{P}_k ranges over the prime ideals of k above the rational prime 2 and χ denotes the quadratic character associated with the extension K/k . \square

Theorem 26. *If the ideal class group of a non-normal quartic CM-field K is of exponent ≤ 2 , then $h_K^- \leq 2^{15}$ and $d_K/d_k \leq 3 \cdot 10^{16}$.*

Proof. We assume that $d_K/d_k \geq 3 \cdot 10^9$, which implies $\rho_K = d_K^{1/4} \geq (d_K/d_k)^{1/4} \geq 222$. Let t denote the number of rational primes ramified in k/\mathbb{Q} and let T be the number of prime ideals ramified in K/k . Let $p_1 = 3 \leq p_2 = 3 < p_3 = 5 \leq p_4 = 5 < p_5 = 7 \dots$ be the nondecreasing sequence of all the odd primes, each one being repeated twice and set $\delta_r = \prod_{k=1}^r p_k$. In the same way, set $\tilde{p}_1 = 3 \leq \tilde{p}_2 = 3 < \tilde{p}_3 = 4 \leq \tilde{p}_4 = 4 < \tilde{p}_5 = 5 \dots$ (where for $k \geq 5$ we set $\tilde{p}_k = p_{k-2}$) and set $\tilde{\delta}_r = \prod_{k=1}^r \tilde{p}_k$. If 2 is not ramified in K , then $d_K/d_k \geq \delta_{t+T}$, whereas if 2 is ramified in K , then $d_K/d_k \geq \tilde{\delta}_{t+T}$. Now, assume that the ideal class group of a non-normal quartic CM-field K is of exponent ≤ 2 . Then $h_K^- \leq 2^{t+T-2}$ (see [LYK, Corollary 17]). Now there are two cases to consider.

First, assume that 2 is not ramified in K . Using the lower bound (42) (which is an increasing function of d_K/d_k), we obtain

$$2^{t+T-2} \geq h_K^- \geq \frac{\sqrt{\delta_{t+T}}}{C_K(\log(\delta_{t+T}) + 3)^2} \quad \text{with } C_K = 9\sqrt{e}\pi^2/32,$$

which implies $t + T \leq 16$, $h_K^- \leq 2^{14}$ and $d_K/d_k \leq 1.5 \cdot 10^{16}$, by using (42).

Second, assume that 2 is ramified in K . Using the lower bound (42) (which is an increasing function of d_K/d_k), we obtain

$$2^{t+T-2} \geq h_K^- \geq \frac{\sqrt{\tilde{\delta}_{t+T}}}{C_K(\log(\tilde{\delta}_{t+T}) + 3)^2} \quad \text{with } C_K = 3\sqrt{e}\pi^2/16,$$

which implies $t + T \leq 17$, $h_K^- \leq 2^{15}$ and $d_K/d_k \leq 2.8 \cdot 10^{16}$, by using (42). \square

Remarks 27.

(1) If we use (13) with $c = 2$ (see Theorem 11) and Lemma 24 we obtain the following lower bounds for relative class numbers of non-normal quartic CM-fields:

$$h_K^- \geq \frac{\sqrt{d_K/d_{K+}}}{C_K(\log(d_K/d_{K+}) + 5\kappa_1/4)^2} \quad \text{where } C_K = \sqrt{e}\pi^2/2 = 8.13611\dots$$

Using this lower bound, we would only obtain that if the ideal class group of a non-normal quartic CM-field K is of exponent ≤ 2 , then $h_K^- \leq 2^{16}$ and $d_K/d_{K+} \leq 9 \cdot 10^{17}$, a 30-fold less satisfactory bound than the previous one.

(2) If we had used (39), we would only have obtained that if the ideal class group of a non-normal quartic CM-field K is of exponent ≤ 2 , then $h_K^- \leq 2^{17}$ and $d_K/d_k \leq 10^{19}$, a 333-fold less satisfactory bound than the previous one (and in fact a bound of no practical use).

(3) The desire to determine all the non-normal quartic CM-fields and all the dihedral octic CM-fields with ideal class groups of exponents ≤ 2 has been a continuous incentive to obtain here as good as possible lower bounds for relative class numbers of CM-fields. These determinations have now been completed and can be found in [LYK].

REFERENCES

- [Arn] S. Arno, The imaginary quadratic fields of class number 4, *Acta Arith.* **60** (1992), 321–324. MR **93b**:11144
- [Bak1] A. Baker, A remark on the class number of quadratic fields, *Bull. London Math. Soc.* **1** (1966), 98–102. MR **39**:2723
- [Bak2] A. Baker, Imaginary quadratic fields of class number 2, *Ann. of Math.* **94** (1971), 139–152. MR **45**:8631
- [Bes] S. Bessassi, Bounds for the degrees of CM-fields of class number one, *Acta Arith.* **106** (2003), 213–245.
- [BP] E. Brown and C. J. Parry, The imaginary bicyclic biquadratic fields with class number 1, *J. Reine Angew. Math.* **266** (1974), 118–120. MR **49**:4974
- [FM] R. Foote and V. K. Murty, Zeros and poles of Artin L -series, *Math. Proc. Cambridge Philos. Soc.* **105** (1989), 5–11. MR **89k**:11109
- [HaHu] K. Hardy and R. H. Hudson, Determination of all imaginary cyclic quartic fields with class number 2, *Trans. Amer. Math. Soc.* **311** (1989), 1–55. MR **89f**:11148
- [HH] K. Horie and M. Horie, CM-fields and exponents of their ideal class groups, *Acta Arith.* **55** (1990), 157–170. MR **91k**:11098
- [HJ] J. Hoffstein and N. Jachnowitz, On Artin's conjecture and the class number of certain CM-fields, I and II, *Duke Math. J* **59** (1989), 553–563 and 565–584. MR **90h**:11104
- [Hof] J. Hoffstein, Some analytic bounds for zeta functions and class numbers, *Invent. Math.* **55** (1979), 37–47. MR **80k**:12019
- [Lan] S. Lang, *Algebraic Number Theory*, Springer-Verlag, Graduate Texts in Math. **110**, Second Edition, 1994. MR **95f**:11085
- [LLO] F. Lemmermeyer, S. Louboutin and R. Okazaki, The class number one problem for some non-abelian normal CM-fields of degree 24, *J. Théorie des Nombres de Bordeaux* **11** (1999), 387–406. MR **2001j**:11104
- [LO] S. Louboutin and R. Okazaki, The class number one problem for some non-abelian normal CM-fields of 2-power degrees, *Proc. London Math. Soc.* (3) **76** (1998), 523–548. MR **99c**:11138
- [Lou1] S. Louboutin, Majorations explicites de $|L(1, \chi)|$, *C. R. Acad. Sci. Paris* **316** (1993), 11–14. MR **93m**:11084
- [Lou2] S. Louboutin, Lower bounds for relative class numbers of CM-fields, *Proc. Amer. Math. Soc.* **120** (1994), 425–434. MR **94d**:11089
- [Lou3] S. Louboutin, Determination of all quaternion octic CM-fields with class number 2, *J. London Math. Soc.* **54** (1996), 227–238. MR **97g**:11122
- [Lou4] S. Louboutin, A finiteness theorem for imaginary abelian number fields, *Manuscripta Math.* **91** (1996), 343–352. MR **97f**:11089
- [Lou5] S. Louboutin, The class number one problem for the non-abelian normal CM-fields of degree 16, *Acta Arith.* **82** (1997), 173–196. MR **98j**:11097
- [Lou6] S. Louboutin, Majorations explicites du résidu au point 1 des fonctions zêta des corps de nombres, *J. Math. Soc. Japan* **50** (1998), 57–69. MR **99a**:11131
- [Lou7] S. Louboutin, Explicit bounds for residues of Dedekind zeta functions, values of L -functions at $s = 1$, and relative class numbers, *J. Number Theory* **85** (2000), 263–282. MR **2002i**:11111
- [Lou8] S. Louboutin, Explicit upper bounds for residues of Dedekind zeta functions and values of L -functions at $s = 1$, and explicit lower bounds for relative class numbers of CM-fields, *Canad. J. Math.* **53** (2001), 1194–1222. MR **2003d**:11167
- [Lou9] S. Louboutin, Majorations explicites de $|L(1, \chi)|$ (quatrième partie), *C. R. Acad. Sci. Paris* **334** (2002), 625–628.
- [LPP] H. W. Lenstra, J. Pila and C. Pomerance, A hyperelliptic smoothness test, II, *Proc. London Math. Soc.* **84** (2002), 105–146.
- [LYK] S. Louboutin, Y.-S. Yang and S.-H. Kwon, The non-normal quartic CM-fields and the dihedral octic CM-fields with ideal class groups of exponent ≤ 2 , *Preprint* (2000).
- [Mur1] M. Ram Murty, An analogue of Artin's conjecture for abelian extensions, *J. Number Theory* **18** (1984), 241–248. MR **85j**:11161
- [Mur2] V. K. Murty, Class numbers of CM-fields with solvable normal closure, *Compositio Math.* **127** (2001), 273–287. MR **2003a**:11147

- [MW] H. L. Montgomery and P. J. Weinberger, Note on small class numbers, *Acta Arith.* **24** (1974), 529–542. MR **50**:9841
- [Od1] A. Odlyzko, Some analytic estimates of class numbers and discriminants, *Invent. Math.* **29** (1975), 275–286. MR **51**:12788
- [Oes] J. Oesterlé, Nombres de classes des corps quadratiques imaginaires, *Séminaire Bourbaki*, 36e année, 1983–84, exposé **631**. Astérisque **121-122** (1985), 309–323. MR **86k**:11064
- [Oka] R. Okazaki, Inclusion of CM-fields and divisibility of relative class numbers, *Acta Arith.* **92** (2000), 319–338. MR **2001h**:11138
- [Pin1] J. Pintz, On Siegel's theorem, *Acta Arith.* **24** (1974), 543–551. MR **49**:2595
- [Pin2] J. Pintz, Elementary methods in the theory of L -functions, VIII, Real zeros of real L -functions, *Acta Arith.* **33** (1977), 89–98. MR **58**:5551g
- [Ram] O. Ramaré, Approximate formulae for $L(1, \chi)$, *Acta Arith.* **100** (2001), 245–256. MR **2002k**:11144
- [Sie] C. L. Siegel, Über die Classenzahl quadratischer Zahlkörper, *Acta Arith.* **1** (1935), 83–86.
- [Sta1] H. M. Stark, A complete determination of the complex quadratic fields of class number 1, *Michigan Math. J.* **14** (1967), 1–27. MR **36**:5102
- [Sta2] H. M. Stark, On complex quadratic fields with class number two, *Math. Comp.* **29** (1975), 289–302. MR **51**:5548
- [Sta3] H. M. Stark, Some effective cases of the Brauer-Siegel Theorem, *Invent. Math.* **23** (1974), 135–152. MR **49**:7218
- [Uch] K. Uchida, Imaginary abelian number fields with class number one, *Tôhoku Math. J.* **24** (1972), 487–499. MR **48**:269
- [Was] L. C. Washington, *Introduction to Cyclotomic Fields*, Springer-Verlag, Graduate Texts in Math. **83**, Second Edition, 1997. MR **97h**:11130
- [YK] H.-S. Yang and S.-H. Kwon, The non-normal quartic CM-fields and the octic dihedral CM-fields with relative class number two, *J. Number Th.* **79** (1999), 175–193. MR **2000h**:11117

INSTITUT DE MATHÉMATIQUES DE LUMINY, UPR 9016, 163 AVENUE DE LUMINY, CASE 907,
13288 MARSEILLE CEDEX 9, FRANCE

E-mail address: loubouti@iml.univ-mrs.fr

PRIMITIVE FREE CUBICS WITH SPECIFIED NORM AND TRACE

SOPHIE HUCZYNSKA AND STEPHEN D. COHEN

ABSTRACT. The existence of a primitive free (normal) cubic $x^3 - ax^2 + cx - b$ over a finite field F with arbitrary specified values of a ($\neq 0$) and b (primitive) is guaranteed. This is the most delicate case of a general existence theorem whose proof is thereby completed.

1. INTRODUCTION

Given q , a power of a prime p , let F denote the finite field $\text{GF}(q)$ of order q and, for a given positive integer n , let E denote its extension $\text{GF}(q^n)$ of degree n . A *primitive element* of E is a generator of the cyclic group E^* . The extension E is also cyclic when viewed as an FG -module, G being the Galois group of E over F , and a generator is called a *free element* of E over F . The core result linking additive and multiplicative structure — the *primitive normal basis theorem* — is that there exists $\alpha \in E$, simultaneously primitive and free over F . Existence of such an element for every extension was demonstrated by Lenstra and Schoof [LeSc] (completing work by Carlitz ([Ca1], [Ca2]) and Davenport [Da]). A computer-free proof of the primitive normal basis theorem is given in [CoHu1].

It is natural to ask whether the result of the Primitive Normal Basis Theorem can be extended by imposing additional conditions on the primitive free element. In particular, we may wish to prescribe the norm or trace of a primitive free element, equivalent to specifying the constant term or the coefficient of x^{n-1} of the corresponding primitive free polynomial. In [CoHa1], it was shown that, given an arbitrary nonzero element $a \in F$, there exists a primitive element ω of E , free over F , such that ω has (E, F) -trace a in F , i.e., $\text{Tr}_{E/F}(\omega) := \sum_{i=0}^{n-1} \omega^{q^i} = a$. Furthermore, in [CoHa2] it was shown that, given an arbitrary primitive element b of F , there exists a primitive element ω of E , free over F , with (E, F) -norm b in F , i.e., $N_{E/F}(\omega) := \prod_{i=0}^{n-1} \omega^{q^i} = \omega^{\frac{q^n-1}{q-1}} = b$.

In [CoHa2], Cohen and Hachenberger posed the following question, known as the PFNT problem. (A similar description of the above problems would be as PFT, PFN respectively, and later we refer to the analogous PNT problem.)

Problem 1.1. Given a finite extension E/F of Galois fields, a primitive element b in F and a nonzero element a in F , does there exist a primitive element $w \in E$, free over F , whose (E, F) -norm and trace equal b and a respectively? Equivalently, amongst all polynomials $\sum_{i=0}^n c_i x^i$ ($c_i \in F$) of degree n over F , does there exist

Received by the editors September 26, 2002 and, in revised form, January 30, 2003.

2000 *Mathematics Subject Classification*. Primary 11T06; Secondary 11A25, 11T24, 11T30.

one that is primitive and free, with $c_{n-1} = -a$ and $c_0 = (-1)^n b$? If so for each pair (a, b) , then the pair (q, n) corresponding to E/F is called a PFNT pair.

Observe that the problem is meaningful only for $n \geq 3$. Clearly the strongest results (and correspondingly those most challenging to prove) occur for small n , since the corresponding polynomials have fewest “degrees of freedom”. The PFNT problem was resolved for all $n \geq 5$ in [Co] (Theorem 1.1); it was observed that the $n = 4$ case was delicate while the $n = 3$ case might prove entirely intractable. The $n = 4$ case was solved in [CoHu2], using a modified version of the $n \geq 5$ approach.

In what follows, we solve the PFNT problem in the affirmative for $n = 3$. Expressing the result in terms of polynomials, we show that: for any prime power q , given $a, b \in F^*$ (b primitive), at least one of the q cubic polynomials $x^3 - ax^2 + cx - b$ ($c \in F$) is primitive and free. Perhaps surprisingly, there are no exceptions.

We have therefore completed the final stage in solving the general PFNT problem, i.e., we have established the existence of a primitive free element with prescribed norm and trace for every extension. The result is summarized in the following theorem.

Theorem 1.2. *Let q be a prime power and $n \geq 3$ an integer. Then (q, n) is a PFNT pair.*

The basic technique ([CoHa2]) of expressing the number of elements with the desired properties in terms of Gauss sums over E yields, if applied directly, estimates in terms of the numbers of prime factors of $q^n - 1$ and irreducible factors of $x^n - 1$. This establishes the result for large n but is inadequate when n is small. In [Co], use of a sieve on both the additive and multiplicative parts produces an expression in terms of the numbers of prime (respectively, irreducible) factors of divisors of $q^n - 1$ (respectively, $x^n - 1$), which are estimated as previously; this approach is more successful in dealing with small n but remains inappropriate for $n < 5$. In this paper, we exploit the idiosyncrasies of the situation when $n = 3$ (allowing us to reduce the PFNT problem to the simpler PNT problem in some cases) and, crucially, employ “external” results to estimate appropriate quantities (i.e., we no longer depend exclusively on the estimates derived from the initial Gauss sum formulation). The extreme delicacy of the $n = 3$ case means that the reductions and improvements which we apply to the basic technique are not merely conveniences, but are vital in establishing the result. Finally, a number of values of $q \leq 256$ (34 in all) had to be checked computationally.

2. PRELIMINARIES

We begin by making some reductions to the problem and formulating the basic theory. The account will be as self-contained as possible, but to avoid excessive repetition, reference will be made to earlier work where appropriate.

By Proposition 4.1 of [CoHa2], (q, n) is a PFNT pair whenever $q - 1$ divides n (so we may assume that $q \neq 2, 4$, in the case when $n = 3$).

From now on, suppose that $a, b \in F$, with $a \neq 0$ and b a primitive element, are given.

Let $m = m(q, n)$ be the greatest divisor of $q^n - 1$ that is relatively prime to $q - 1$ (so in particular m divides $\frac{q^n - 1}{(q - 1)(n, q - 1)}$). In [Co] it was demonstrated that, if $w \in E$ has (E, F) -norm b , then to guarantee that w is primitive it suffices to show that w is m -free in E (i.e., that $w = v^d$, where $v \in E$ and $d|m$, implies $d = 1$).

Analogously for the additive part, let $M = M(q, n)$ be the monic divisor of $x^n - 1$ (over F) of maximal degree that is prime to $x - 1$. So $M = \frac{x^n - 1}{x^{p^l} - 1}$ where $n = n_0 p^l$, $p = \text{char} F$ and $p \nmid n_0$. It was shown in [Co] that, if $w \in E$ has (nonzero) (E, F) -trace a , then to guarantee that w is free over F it suffices to show that w is M -free in E (i.e., that $w = h^\sigma(v)$, where $v \in E$ and h is an F -divisor of M , implies $h = 1$).

Define $N(t, T)$ to be the number of elements of E that

- (i) are t -free ($t \in \mathbb{Z}$, $t|m$),
- (ii) are T -free ($T(x) \in F[x]$, $T|x^n - 1$),
- (iii) have norm b , and
- (iv) have trace a .

Write $\pi(t, T)$ for $q(q - 1)N(t, T)$. Assume throughout that $t|m$ and $T|x^n - 1$.

Next, we express the characteristic functions of the four subsets of E (or E^*) defined by the conditions (i)-(iv) in terms of characters on E or F .

I. *The set of $w \in E^*$ with $N_{E/F}(w) = b$.* The characteristic function of the subset of E^* comprising elements with norm b is

$$\frac{1}{q-1} \sum_{\nu \in \hat{F}^*} \nu(N(w)b^{-1}),$$

where \hat{F}^* denotes the group of multiplicative characters of F^* , and the norm $N_{E/F}$ is abbreviated to N .

II. *The set of $w \in E^*$ with $\text{Tr}_{E/F}(w) = a$.* The characteristic function of the subset of E comprising elements with trace a is

$$\frac{1}{q} \sum_{c \in F} \lambda(c(T(w) - a)),$$

where λ is the canonical additive character of F and the trace $\text{Tr}_{E/F}$ is abbreviated to T .

III. *The set of $w \in E^*$ that are t -free.* The characteristic function for the subset of t -free elements ($t|m$) of E^* is

$$\theta(t) \int_{d|t} \eta_d(w), \quad w \in E^*,$$

where $\theta(t) = \frac{\phi(t)}{t}$, η_d denotes a character of order d ($d|m$) in \hat{E}^* and, using the notation introduced in [Co], the integral notation is shorthand for a weighted sum:

$$\int_{d|t} \eta_d := \sum_{d|t} \frac{\mu(d)}{\phi(d)} \sum_{(d)} \eta_d.$$

IV. *The set of $w \in E$ that are T -free over F .* The characteristic function of the set of T -free elements of E takes the form

$$\Theta(T) \int_{D|T} \chi_{\delta_D}(w), \quad w \in E,$$

where $\Theta(T) = \frac{\Phi(T)}{T}$, χ is the canonical additive character on E and, as defined in [Co], $\{\chi_{\delta_D} : \delta_D \in \Delta_D\}$ (where $\chi_{\delta}(w) := \chi(\delta w)$, $w \in E$) is the set of all additive

characters of E of order D ($D|x^n - 1$). Again, the integral notation represents a weighted sum:

$$\int_{D|g} \chi_{\delta_D} := \sum_{D|g} \frac{\mu(D)}{\Phi(D)} \sum_{(\delta_D)} \chi_{\delta_D}.$$

Using these characteristic functions, we derive an expression for $\pi(t, T)$:

$$(2.1) \quad \pi(t, T) = \theta(t)\Theta(T) \int_{d|t} \int_{D|T} \sum_{\nu \in \hat{F}^*} \sum_{c \in F} \bar{\nu}(b)\bar{\lambda}(ac) \sum_{w \in E} (\eta_d \tilde{\nu}(w) \chi((\delta_D + c)w))$$

where $\tilde{\nu}(w) = \nu(N(w))$ and $\chi(cw) = \lambda(cT(w))$ (cf. [Co], equation (2.2)).

We shall now specialise to the case when $n = 3$. Observe that, if $p|n$ (i.e., if $q = 3^k$ for some $k \in \mathbb{N}$), then $M = 1$ and the PFNT problem reduces to the (nonzero) PNT problem. If $q \equiv 2 \pmod{3}$, then $M = x^2 + x + 1$ is irreducible over F ; by Lemma 3.5 of [Co], $\pi(m, M) > 0$ if and only if $\pi(m, 1) > 0$, and so the PFNT problem reduces to the (nonzero) PNT problem in this case also. Hence only in the case when $q \equiv 1 \pmod{3}$ need the full PFNT problem be considered. When $q \equiv 1 \pmod{3}$, $M = (x - \gamma)(x - \gamma^2)$ (where $\gamma \in F$ is such that $\gamma^3 = 1$, $\gamma \neq 1$).

With regard to the multiplicative part of the problem, we note that all prime divisors of m must be congruent to 1 modulo 6. For, since $m|(q^2 + q + 1)$, an odd number, then m is odd. Furthermore, suppose that for some prime l , $l|m$. Then $l|q^3 - 1$ but $l \nmid q - 1$; hence $\text{ord}_l q = 3$. By Fermat's Little Theorem, $q^{l-1} \equiv 1 \pmod{l}$ since $l \nmid q$. So $3|l - 1$, i.e., $l \equiv 1 \pmod{3}$. Thus all prime divisors of m lie in the set $\{7, 13, 19, 31, 37, \dots\}$.

Our strategy for proving the PFNT problem for $n = 3$ is to apply a sieving technique. We shall use the basic sieving inequality introduced in Theorem 3.1 of [Co]. Let $d|m$ and $f|x^n - 1$. Then (d_i, f_i) ($i = 1, \dots, r$ for $r \in \mathbb{N}$) will be called *complementary divisor pairs* with *common divisor pair* (d_0, f_0) if the primes in $\text{lcm}\{d_1, \dots, d_r\}$ are precisely those in d , the irreducibles in $\text{lcm}\{f_1, \dots, f_r\}$ are precisely those in f , and for any distinct pair (i, j) , the primes and irreducibles in $\text{gcd}(d_i, d_j)$ and $\text{gcd}(f_i, f_j)$ are precisely those in d_0 and f_0 respectively. Observe that the value of $\pi(d, f)$ will depend only on which ‘‘atoms’’ (primes/irreducibles) are present in d and f , not on the power to which the atoms occur.

Lemma 2.1 (Sieving inequality). *For divisors d of m and f of $x^n - 1$, let $\{(d_1, f_1), \dots, (d_r, f_r)\}$ be complementary divisor pairs of (d, f) with common divisor (d_0, f_0) . Then*

$$(2.2) \quad \pi(d, f) \geq \left(\sum_{i=1}^r \pi(d_i, f_i) \right) - (r - 1)\pi(d_0, f_0).$$

In the PNT case, where there is no additive component, the sieve will clearly take the following simpler form. For divisors d of m , let d_1, \dots, d_r be divisors of d (with common divisor d_0) such that the primes in $\text{lcm}\{d_1, \dots, d_r\}$ are precisely those in d and, for any distinct pair (i, j) , the primes in $\text{gcd}(d_i, d_j)$ are precisely those in d_0 . Then

$$(2.3) \quad \pi(d, 1) \geq \left(\sum_{i=1}^r \pi(d_i, 1) \right) - (r - 1)\pi(d_0, 1).$$

In the next section, we establish estimates for $\pi(t, 1)$ ($t|m$).

3. ESTIMATES FOR INTEGER FACTORS

In this section we obtain new estimates for the number $N(t, 1)$ of t -free elements of E with prescribed norm and trace, where $t \in \mathbb{N}$ is a divisor of m . We improve upon the estimates of [Co] by applying some deep results of Katz arising from the study of Soto-Andrade sums [Ka].

Lemma 3.1 ([Ka], Theorem 4). *Suppose that $n \geq 2$. Then*

$$(3.1) \quad \left| N(1, 1) - \frac{q^n - 1}{q(q - 1)} \right| \leq nq^{\frac{n-2}{2}},$$

i.e.,

$$(3.2) \quad |\pi(1, 1) - (q^n - 1)| \leq n \left(1 - \frac{1}{q} \right) q^{\frac{n+2}{2}}.$$

In particular, for $n = 3$, Lemma 3.1 has the form

$$|\pi(1, 1) - (q^3 - 1)| \leq 3 \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}}.$$

Note that this is an improvement, by a factor of approximately $\frac{q^{\frac{1}{2}}}{3}$, on the estimate

$$|\pi(1, 1) - q^3| \leq \left(1 - \frac{(e + 1)}{q} \right) q^3$$

($e := \gcd(3, q - 1)$), obtainable from Corollary 2.2 of [Co] but useless as a lower bound. It is such increases in accuracy that allow us to solve the $n = 3$ case where the method of [Co] fails.

Next, we estimate $N(t, 1)$ where $t|m$, $t > 1$.

Lemma 3.2 ([Ka], Corollary of Theorem 3 bis). *Let η be a character of E of order d , where $d|m$, $d > 1$. Suppose $\eta^{n(q-1)}$ is not trivial. Set*

$$M(\eta) = \sum_{\substack{x \in E \\ N(x)=b \\ T(x)=a}} \eta(x).$$

Then

$$|M(\eta)| \leq nq^{\frac{n-2}{2}}.$$

This lemma is applicable when $n = 3$ to all $\eta_d \in \hat{F}^*$ ($d|m$, $d > 1$). For, consider some $\eta \in \hat{F}^*$ of order d , where $d|m$ and $d > 1$. Clearly η^{q-1} cannot be trivial or have order 3, since $(d, q - 1) = 1$ and $(d, 3) = 1$.

Corollary 3.3. *Let $t|m$, $t > 1$, and $t_0|t$, $t_0 \geq 1$. Then*

$$(3.3) \quad \left| \pi(t, 1) - \frac{\theta(t)}{\theta(t_0)} \pi(t_0, 1) \right| \leq n\theta(t)(W(t) - W(t_0)) \left(1 - \frac{1}{q} \right) q^{\frac{n+2}{2}}.$$

Proof. By definition,

$$N(t, 1) = \theta(t) \sum_{\substack{w \in E \\ N(w)=b \\ T(w)=a}} \int_{d|t} \eta_d(w) = \theta(t) \int_{d|t} M(\eta_d),$$

and so

$$N(t, 1) - \frac{\theta(t)}{\theta(t_0)} N(t_0, 1) = \theta(t) \int_{d|t_0}^{d|t} M(\eta_d).$$

By Lemma 3.2,

$$\left| N(t, 1) - \frac{\theta(t)}{\theta(t_0)} N(t_0, 1) \right| \leq \theta(t) (W(t) - W(t_0)) n q^{\frac{n-2}{2}},$$

and hence

$$\left| \pi(t, 1) - \frac{\theta(t)}{\theta(t_0)} \pi(t_0, 1) \right| \leq n \theta(t) (W(t) - W(t_0)) \left(1 - \frac{1}{q} \right) q^{\frac{n+2}{2}}.$$

In particular, for $n = 3$,

$$(3.4) \quad \left| \pi(t, 1) - \frac{\theta(t)}{\theta(t_0)} \pi(t_0, 1) \right| \leq 3 \theta(t) (W(t) - W(t_0)) \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}}.$$

□

4. THE (NONZERO) PNT PROBLEM

Recall from Section 2 that, if q is a power of 3 or if $q \equiv 2 \pmod{3}$, then the PFNT problem reduces to the (nonzero) PNT problem ("nonzero" refers to the fact that the prescribed trace a is nonzero). Hence, to establish the result in these cases, it suffices to show that $\pi(m, 1) > 0$.

In order to simplify notation, from this point onwards we shall adopt the convention that all unmarked summation signs have index i running from $i = 1$ to s (where s is the number of distinct primes dividing m), and that $p[i]$ denotes the i th prime congruent to 1 modulo 6, i.e., the i th element of the set $\{7, 13, 19, 31, 37, \dots\}$.

The following lemma provides a useful upper bound for $W(t)$.

Lemma 4.1. *For any positive integer t ,*

$$(4.1) \quad W(t) \leq c_t t^{1/6},$$

where $c_t = \frac{2^r}{(p_1 \dots p_r)^{1/6}}$, and p_1, \dots, p_r are the distinct primes less than 64 that divide t . In particular, if $p_i \equiv 1 \pmod{6}$ for all $i = 1, \dots, r$, then $c_t < 3.08$.

(The proof is obvious using multiplicativity.)

Proposition 4.2. *Suppose q is a prime power, $q \not\equiv 1 \pmod{3}$. Then $(q, 3)$ is a PNT pair for all $q \geq 622,346$. In particular, $(3^k, 3)$ is a PNT pair for all $k \in \mathbb{N}$, $k > 12$.*

Proof. Apply the bounds of Lemma 3.1 and Corollary 3.3 directly, without sieving. Then

$$\pi(m, 1) \geq \theta(m) \left\{ (q^3 - 1) - 3 \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}} \right\} - 3 \theta(m) (W(m) - 1) \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}},$$

and so $\pi(m, 1) > 0$ whenever

$$(4.2) \quad q^{\frac{1}{2}} > 3W(m) \left(1 - \frac{1}{q} \right) + \frac{1}{q^{\frac{5}{2}}}.$$

Using the approximation of Lemma 4.1 for $W(m)$, $(q, 3)$ is a PNT pair whenever

$$(4.3) \quad q > 3c_m(q-1)^{\frac{5}{6}} + \frac{1}{q^2},$$

where $c_m = 3.08$. This inequality holds for integers $q \geq 622,346$, and so establishes the result. \square

The following simplification applies in the case when $3|q$ and m is prime.

Lemma 4.3. *Let $q = 3^k$, $k \in \mathbb{N}$, so that $m = q^2 + q + 1$. Suppose that m is prime. Then*

$$N(m, 1) = N(1, 1),$$

where $N(t, 1)$ ($t|m$) is the number of t -free elements of E with trace and norm equal to a and b respectively ($a, b \in F$, $a \neq 0$, b primitive).

Proof. Suppose $\alpha \in E$ (i.e., trivially 1-free) with $\text{Tr}(\alpha) = a$, $N(\alpha) = b$, but $\alpha = \beta^m$. Then $\alpha^{q-1} = 1$, i.e., $\alpha \in \text{GF}(q)$. Hence, $\text{Tr}_{E/F}(\alpha) = 3\alpha$, which equals 0 since $\text{char} F = 3$, a contradiction since $a \neq 0$. \square

Proposition 4.4. *Suppose q is a prime power, $q \not\equiv 1 \pmod{3}$, and let $m = p_1^{\alpha_1} \cdots p_s^{\alpha_s}$. Then $(q, 3)$ is a PNT pair whenever*

$$(4.4) \quad \pi(1, 1) \left\{ 1 - \sum \frac{1}{p_i} \right\} - 3 \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}} \sum \left(1 - \frac{1}{p_i} \right) > 0,$$

and so certainly whenever

$$(4.5) \quad q^{\frac{1}{2}} > C_s$$

where

$$C_s := 3 \left(2 + \frac{s-1}{1 - \sum_{i=1}^s \frac{1}{p[i]}} \right) + \frac{1}{3^{\frac{5}{2}}},$$

where $p[i]$ is the i th prime congruent to 1 modulo 6.

Proof. Apply the sieve with atomic divisors. Using the results of Corollary 3.3, $\pi(m, 1) > 0$ whenever

$$\pi(1, 1) \left\{ 1 - \sum \frac{1}{p_i} \right\} - 3 \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}} \sum \left(1 - \frac{1}{p_i} \right) > 0.$$

By Lemma 3.1, $\pi(m, 1) > 0$ if

$$(4.6) \quad q^{\frac{1}{2}} > 3 \left(1 - \frac{1}{q} \right) \left(1 + \frac{\sum (1 - \frac{1}{p_i})}{1 - \sum \frac{1}{p_i}} \right) + \frac{1}{q^{\frac{5}{2}}}.$$

Replacing the right-hand side of (4.6) by a larger quantity depending solely on s , we see that the desired result certainly holds when

$$(4.7) \quad q^{\frac{1}{2}} > C_s$$

where

$$C_s := 3 \left(2 + \frac{s-1}{1 - \sum \frac{1}{p[i]}} \right) + \frac{1}{3^{\frac{5}{2}}}.$$

Observe that, since C_s is a constant for fixed s and increases as s increases (for all s such that $\sum_{i=1}^s \frac{1}{p[i]} < 1$), $q^{\frac{1}{2}} > C_{s_1}$ for some s_1 implies that $q^{\frac{1}{2}} > C_s$ for all $s \leq s_1$. \square

Proposition 4.5. (i) Suppose $q = 3^k$ ($k \in \mathbb{N}$, $k \geq 5$ or $k = 3$). Then $(q, 3)$ is a PFNT pair.

(ii) Suppose $q \equiv 2 \pmod{3}$ and $q \leq 622,346$ but $q \notin \{5, 8, 11, 17, 23, 29, 32, 47, 53, 107, 137, 149, 191\}$. Then $(q, 3)$ is a PNT pair.

Proof. (i) Lemma 4.2 has established the result for $k > 12$; so we need consider only $k \leq 12$.

Let $m = p_1^{\alpha_1} \dots p_s^{\alpha_s}$. We apply Proposition 4.4. For all $q = 3^k$ with $k \leq 12$ we have $s \leq 5$. Since $C_5^2 < 577 < 3^6$, the result holds for $q = 3^k$, $k \geq 6$. The result is established for $k = 5$ ($s = 2$), since $3^5 > 71 > C_2^2$. When $k = 3$, m ($= 757$) is prime; hence by Lemma 4.3, m may be replaced by 1. Inequality (4.2) is then satisfied, since $\sqrt{27} > 2.8892$.

(ii) For $q > 2$, let $m = p_1^{\alpha_1} \dots p_s^{\alpha_s}$; since $m \leq q^2 + q + 1$, $s \leq 8$ for $q < 622,346$ (merely by size considerations). As in part (i), we apply Proposition 4.4.

Since inequality (4.5) holds for all relevant $q > 1622$, the result is established for prime powers $q \geq 1637$. For $q < 1622$, we find that $s \leq 4$; then the desired result holds for $q > 361$, i.e., for all $q \geq 367$. Since the smallest $q \equiv 2 \pmod{3}$ with $s = 4$ is $q = 809$, use of inequality (4.5) with $s = 3$ then establishes the result for $q > 204$, i.e., $q \geq 227$. However, even use of exact values fails for those $q < 204$ with $s = 3$, namely $\{107, 137, 149, 191\}$. Similarly, (4.5) holds with $s = 2$ for $q > 98$, and thus establishes the result for all $q \geq 101$ (apart from the preceding exceptions). Use of exact values in inequality (4.6) yields the result for $q = 83$ ($m = 19 \cdot 367$, $\sqrt{83} > 9.110 > 9.065 > \text{right side of (4.6)}$). Values of q with $s = 2$ for which exact values are insufficient are $\{11, 23, 29, 32, 47, 53\}$. Finally, $q^{\frac{1}{2}} > C_1$ for all $q > 36$, i.e., $q \geq 41$, which establishes all remaining cases with the exception of $\{5, 8, 17\}$. \square

5. THE PFNT PROBLEM

In this section, the full PFNT problem will be solved for the case when $q \equiv 1 \pmod{3}$.

Denote by L a linear factor of $M (= x^2 + x + 1)$; L may take the values $x - \gamma$ or $x - \gamma^2$, where $\gamma \in F$ is such that $\gamma^3 = 1$, $\gamma \neq 1$. We begin by deriving estimates for the number $N(1, L)$ of L -free elements of E with prescribed norm and trace. For economy of calculation, it is in fact desirable to consider the difference between $\pi(1, L)$ and $\theta(L)\pi(1, 1)$ (in some sense the "error term"). We will prove the following lemma.

Lemma 5.1. Let $q \equiv 1 \pmod{3}$. Then

$$(5.1) \quad \left| \pi(1, x - \gamma) + \pi(1, x - \gamma^2) - 2 \left(1 - \frac{1}{q} \right) \pi(1, 1) \right| \leq 2q^{\frac{5}{2}} \left(1 - \frac{3}{q} - \frac{2}{q^2} \right) + 2q^2 \left(1 - \frac{3}{q} \right).$$

First, we establish some results about δ_L . For a polynomial $f(x)$, let f^σ denote the polynomial obtained from f by replacing x^i by x^{q^i} .

Lemma 5.2. If $D|x^{n/k} - 1$ (where $k|n$), then δ_D is a root of $(x^{n/k} - 1)^\sigma$, i.e., $\delta_D \in \text{GF}(q^{n/k})$.

Proof. Set $R = q^{n/k}$. So for the canonical character χ_1 of E , $\chi_1(w) = \lambda(\text{Tr}_{R^k/p}(w))$ ($w \in E$), where $\lambda(x) = e^{\frac{2\pi x}{p}}$. Let $\chi(w) = \chi_\delta(w) = \lambda(\text{Tr}_{R^k/p}(\delta w))$ and suppose $\delta \in$

$\text{GF}(R)$; so $\delta^R = \delta$. Then

$$\begin{aligned}\chi(w^R) &= \lambda(\text{Tr}_{R^k/p}(\delta w^R)) \\ &= \lambda(\text{Tr}_{R/p}(\text{Tr}_{R^k/R}(\delta^R w^R))) \\ &= \lambda(\text{Tr}_{R/p}(\text{Tr}_{R^k/R}(\delta w))) \\ &= \lambda(\text{Tr}_{R^k/p}(\delta w)) \\ &= \chi(w).\end{aligned}$$

Hence $\chi(w^R - w) = 1$ for all $w \in E$. So for any $D|x^{n/k} - 1$, i.e., $D^\sigma|x^R - x$, $\chi_\delta(D^\sigma(w)) = 1$. Thus $\delta = \delta_D$ for some $D|x^{n/k} - 1$. Letting δ vary in $\text{GF}(R)$ accounts for all R characters of order dividing $x^{n/k} - 1$. \square

Lemma 5.3. *Suppose $q \equiv 1 \pmod{3}$, and let $\gamma \in \text{GF}(q)$ be such that $\gamma^3 = 1$, $\gamma \neq 1$.*

- (i) *Let $D = x - \gamma$. Then $(x - \gamma^2)^\sigma(\delta_D) = 0$, i.e., $\delta_D^q = \gamma^2 \delta_D$.*
- (ii) *Let $D = x - \gamma^2$. Then $(x - \gamma)^\sigma(\delta_D) = 0$, i.e., $\delta_D^q = \gamma \delta_D$.*

Proof. (i) Suppose $\delta^q = \gamma^2 \delta$. Define $\chi(w) = \chi_1(\delta w) = \lambda(\text{Tr}_{q^3/p}(\delta w))$, $w \in E = \mathbb{F}_{q^3}$. Then

$$\begin{aligned}\chi(w^q - \gamma w) &= \lambda(\text{Tr}_{q/p}[\text{Tr}_{q^3/q}(\delta(w^q - \gamma w))]) \\ &= \lambda(\text{Tr}_{q/p}[\text{Tr}_{q^3/q}(\gamma \delta^q w^q - \gamma \delta w)]) \\ &= \lambda(\text{Tr}_{q/p}[\gamma \text{Tr}_{q^3/q}((\delta w)^q - \delta w)]) \\ &= 1,\end{aligned}$$

since $\text{Tr}_{q^3/q}((\delta w)^q - \delta w) \equiv 0$. So the F -order of χ is $x - \gamma$. This accounts for all characters with F -order $x - \gamma$.

- (ii) Replace γ by γ^2 in (i). \square

We are now ready to prove Lemma 5.1. Throughout this discussion, $G_n(\eta)$ (where η is a multiplicative character on $\mathbb{F}_{q^n}^*$) will denote a Gauss sum in $\mathbb{F}_{q^n}^*$. We will use the notation $J_a(\nu_1, \dots, \nu_k)$ (where $a \in F$, ν_1, \dots, ν_k are multiplicative characters of F , $k \in \mathbb{N}$) to denote the Jacobi sum

$$\sum_{c_1 + \dots + c_k = a} \nu_1(c_1) \dots \nu_k(c_k).$$

For extra background material, the reader may consult texts such as [LiNi].

Proof of Lemma 5.1. By equation (2.1), since $\Theta(L) = (1 - \frac{1}{q})$,

$$\begin{aligned}(5.2) \quad &\pi(1, L) - \Theta(L)\pi(1, 1) \\ &= \Theta(L) \left(-\frac{1}{q-1} \right) \sum_{\nu \in \hat{F}^*} \sum_{c \in F} \sum_{(\delta_L)} \bar{\nu}(b) \bar{\lambda}(ac) \sum_{w \in E} \tilde{\nu}(w) \chi((\delta_L + c)w),\end{aligned}$$

where δ_L runs through all $\Phi(L)$ elements of Δ_L (i.e., χ_{δ_L} runs through all additive characters of E of order L). Separating the term for which $c = 0$, we have

$$(5.3) \quad \begin{aligned} \pi(1, L) - \Theta(L)\pi(1, 1) &= -\frac{1}{q} \left\{ \sum_{\nu \in \hat{F}^*} \sum_{(\delta_L)} \bar{\nu}(b) \sum_{w \in E} \tilde{\nu}(w) \chi(\delta_L w) \right. \\ &\quad \left. + \sum_{\nu \in \hat{F}^*} \sum_{c \in F^*} \sum_{(\delta_L)} \bar{\nu}(b) \bar{\lambda}(ac) \sum_{w \in E} \tilde{\nu}(w) \chi((\delta_L + c)w) \right\}. \end{aligned}$$

For the first term on the right side of (5.3), using the fact that $\delta_L \neq 0$, replace w by $\frac{w}{\delta_L}$ to obtain

$$\sum_{\nu \in \hat{F}^*} \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \sum_{(\delta_L)} \tilde{\nu}(\delta_L).$$

Since $F^* \Delta_D = \Delta_D$,

$$\sum_{(\delta_L)} \tilde{\nu}(\delta_L) = \frac{1}{q-1} \sum_{(\delta_L)} \sum_{c \in F^*} \tilde{\nu}(c\delta_L) = \frac{1}{q-1} \sum_{(\delta_L)} \tilde{\nu}(\delta_L) \left(\sum_{c \in F^*} \tilde{\nu}(c) \right),$$

and the inner sum equals 0 unless $\nu^* (:= \tilde{\nu}|_F)$ is trivial, when it equals $q-1$.

Note that, for $k \in F$, $\nu^*(k) = \tilde{\nu}(k) = \nu(N(k)) = \nu(k^3)$, i.e., $\nu^* = \nu^3$. So the first term of (5.3) can be simplified to

$$\sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 = \nu_1}} \sum_{(\delta_L)} \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \tilde{\nu}(\delta_L).$$

For the second term on the right side of (5.3) (i.e., the part for which $c \neq 0$), replace δ_L by $c\delta_L$, then w by $\frac{w}{c(\delta_L+1)}$, to get

$$(5.4) \quad \sum_{\nu \in \hat{F}^*} \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \sum_{(\delta_L)} \tilde{\nu}(\delta_L + 1) \sum_{c \in F^*} \bar{\lambda}(ac) \tilde{\nu}(c).$$

Consider the inner sum $\sum_{c \in F^*} \bar{\lambda}(ac) \tilde{\nu}(c)$ of (5.4); in the case when $\nu^3 = \nu_1$, this reduces to a sum over additive characters of F , while for $\nu^3 \neq \nu_1$, a Gauss sum over F is obtained. Thus the second term of (5.3) may be expanded as

$$- \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 = \nu_1}} \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \sum_{(\delta_L)} \tilde{\nu}(\delta_L + 1) + \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 \neq \nu_1}} \nu^*(a) \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \tilde{G}_1(\nu^*) \sum_{(\delta_L)} \tilde{\nu}(\delta_L + 1).$$

Hence,

$$\begin{aligned} \pi(1, L) - \Theta(L)\pi(1, 1) &= -\frac{1}{q} \left(\sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 \neq \nu_1}} \nu \left(\frac{a^3}{b} \right) G_3(\tilde{\nu}) \bar{G}_1(\nu^*) \sum_{(\delta_L)} \tilde{\nu}(\delta_L + 1) \right. \\ &\quad \left. + \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 = \nu_1 \\ \tilde{\nu} \neq \eta_1}} \nu \left(\frac{1}{b} \right) G_3(\tilde{\nu}) \sum_{(\delta_L)} (\tilde{\nu}(\delta_L) - \tilde{\nu}(\delta_L + 1)) \right) \\ &= \frac{1}{q} \left(\sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 = \nu_1 \\ \nu \neq \nu_1}} \nu \left(\frac{1}{b} \right) G_1^3(\nu) \sum_{(\delta_L)} [\bar{\nu}(N(\delta_L + 1)) - \bar{\nu}(N(\delta_L))] \right. \\ &\quad \left. - \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 \neq \nu_1}} \sum_{(\delta_L)} \nu \left(\frac{a^3}{b} \right) \bar{\nu}(N(\delta_L + 1)) \bar{G}_1(\nu^3) G_1^3(\nu) \right), \end{aligned}$$

since $G_3(\tilde{\nu}) = G_1^3(\nu)$ by the Davenport-Hasse Theorem (see [LiNi], Chapter 5).

Consider the specific values that may be taken by L , namely $L = x - \gamma$ and $L = x - \gamma^2$. By Lemma 5.2, since these L are divisors of $x^3 - 1$, $\delta_L^{q^3} = \delta_L$. Using Lemma 5.2 and Lemma 5.3, we find that $\delta_L^3 \in \mathbb{F}_q^*$ but $\delta_L \notin \mathbb{F}_q^*$, and so $\delta_L^3 = c$, where c is a non-cube in F . In fact, $\{\delta_{x-\gamma}\} \cup \{\delta_{x-\gamma^2}\} = \{e \in E: e^{3(q-1)} = 1, e^{(q-1)} \neq 1\} = \{\text{cube roots of } c, c \text{ a non-cube in } F\}$, a set of cardinality $2(q-1)$.

In the case when $L = x - \gamma$, using Lemma 5.3 we get

$$N(\delta_L) = \delta_L \delta_L^q \delta_L^{q^2} = \delta_L(\gamma^2 \delta_L)(\gamma \delta_L) = \delta_L^3 = c$$

and

$$N(1 + \delta_L) = (1 + \gamma + \gamma^2)(\delta_L + \delta_L^2) = (1 + \delta_L^3) = 1 + c.$$

The same values are obtained when $L = x - \gamma^2$.

Denote $x - \gamma$ and $x - \gamma^2$ by L_1 and L_2 respectively. Let $\nu_3 \in \hat{F}^*$ be an arbitrary character of degree 3. Then

$$\pi(1, L_1) + \pi(1, L_2) - 2\Theta(L)\pi(1, 1) = \frac{2}{q} \{S_2 - S_1\}$$

where

$$(5.5) \quad S_1 := \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 \neq \nu_1}} \nu \left(\frac{a^3}{b} \right) \bar{G}_1(\nu^3) G_1^3(\nu) \sum_{c \in F^*} \left(1 - \frac{1}{2}(\nu_3(c) + \nu_3^2(c)) \right) \bar{\nu}(1 + c)$$

and

$$(5.6) \quad S_2 := \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 = \nu_1 \\ \nu \neq \nu_1}} \nu \left(\frac{1}{b} \right) G_1^3(\nu) \sum_{c \in F^*} \left[1 - \frac{1}{2}(\nu_3(c) + \nu_3^2(c)) \right] (\bar{\nu}(1+c) - \bar{\nu}(c)).$$

Consider S_1 (as given by (5.5)). It may be written in the form

$$S_1 = \sum_{\substack{\nu \in \hat{F}^* \\ \nu^3 \neq \nu_1}} \nu \left(\frac{a^3}{b} \right) \bar{G}_1(\nu^3) G_1^3(\nu) \sigma_1,$$

say, where $\sigma_1 := \sum_{c \in F^*} (1 - \frac{1}{2}(\nu_3(c) + \nu_3^2(c))) \bar{\nu}(1+c)$. Then

$$\begin{aligned} \sigma_1 &= \sum_{c \in F^*} \bar{\nu}(1+c) - \frac{1}{2} \nu_3(-1) \sum_{c \in F^*} \nu_3(c) \bar{\nu}(1-c) - \frac{1}{2} \nu_3^2(-1) \sum_{c \in F^*} \nu_3^2(c) \bar{\nu}(1-c) \\ &= -1 - \frac{1}{2} (J_1(\nu_3, \bar{\nu}) + J_1(\nu_3^2, \bar{\nu})). \end{aligned}$$

Since each Jacobi sum has absolute value \sqrt{q} ,

$$|S_1| \leq (q-4) \sqrt{q} q^{\frac{3}{2}} (1 + \sqrt{q}),$$

i.e.,

$$(5.7) \quad \frac{2}{q} |S_1| \leq 2q^{\frac{5}{2}} \left(1 - \frac{4}{q} \right) \left(1 + \frac{1}{\sqrt{q}} \right).$$

Now consider S_2 (as given by (5.6)). For a given ν with $\nu^3 = \nu_1$, $\nu \neq \nu_1$, the inner sum σ_2 has the form

$$\sigma_2 := \sum_{c \in F^*} \left(1 - \frac{1}{2}(\nu_3(c) + \nu_3^2(c)) \right) (\bar{\nu}(1+c) - \bar{\nu}(c))$$

where ν_3 is an arbitrary character of order 3. Without loss of generality, we may set $\nu_3 := \nu$ in our expression for σ_2 :

$$\begin{aligned} \sigma_2 &= \sum_{c \in F^*} \bar{\nu}(1+c) - \sum_{c \in F^*} \bar{\nu}(c) \\ &\quad - \frac{1}{2} \left(\sum_{c \in F^*} \nu(c) \bar{\nu}(1+c) - \sum_{c \in F^*} \nu(c) \bar{\nu}(c) \right. \\ &\quad \left. + \sum_{c \in F^*} \nu^2(c) \bar{\nu}(1+c) - \sum_{c \in F^*} \nu^2(c) \bar{\nu}(c) \right) \\ &= (-1) - 0 - \frac{1}{2} (J_1(\nu, \bar{\nu}) - (q-1) + J_1(\nu^2, \bar{\nu}) - 0) \\ &= \frac{1}{2} (q-2) - \frac{1}{2} J_1(\nu^2, \bar{\nu}), \end{aligned}$$

since $\nu \bar{\nu} = \nu_1$ and $\nu^2 \bar{\nu} = \nu$.

Thus $|\sigma_2| \leq \frac{1}{2}(q-2) + \frac{1}{2}\sqrt{q}$. Hence,

$$|S_2| \leq 2q^{\frac{3}{2}} \left(\frac{1}{2}(q-2) + \frac{1}{2}\sqrt{q} \right) = q^{\frac{5}{2}} \left(1 - \frac{2}{q} \right) + q^2,$$

i.e.,

$$(5.8) \quad \frac{2}{q}|S_2| \leq 2q^{\frac{3}{2}} \left(1 - \frac{2}{q}\right) + 2q.$$

Combining inequalities (5.7) and (5.8), we get

$$\begin{aligned} & |\pi(1, L_1) + \pi(1, L_2) - 2\Theta(L)\pi(1, 1)| \\ & \leq 2q^{\frac{5}{2}} \left(1 - \frac{4}{q}\right) \left(1 + \frac{1}{\sqrt{q}}\right) + 2q^{\frac{3}{2}} \left(1 - \frac{2}{q}\right) + 2q \\ & = 2q^{\frac{5}{2}} \left(1 - \frac{3}{q} - \frac{2}{q^2}\right) + 2q^2 \left(1 - \frac{3}{q}\right), \end{aligned}$$

which proves the result. \square

The following is a sufficient condition for $(q, 3)$ to be a PFNT pair.

Lemma 5.4. *Suppose $q \equiv 1 \pmod{3}$. Then $(q, 3)$ is a PFNT pair whenever*

$$(5.9) \quad \begin{aligned} & \pi(1, 1) \left(\theta(m) - \frac{2}{q} \right) \\ & > 3\theta(m)(W(m) - 1) \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}} + 2q^{\frac{5}{2}} \left(1 - \frac{3}{q} - \frac{2}{q^2} \right) + 2q^2 \left(1 - \frac{3}{q} \right). \end{aligned}$$

Proof. Apply the sieve in the following form:

$$\pi(m, M) \geq \pi(m, 1) + \pi(1, x - \gamma) + \pi(1, x - \gamma^2) - 2\pi(1, 1).$$

Using the lower bounds for $\pi(m, 1)$ and the $\pi(1, L_i)$ ($i = 1, 2$) from inequalities (3.3) and Lemma 5.1, we see that $\pi(m, M) > 0$ whenever (5.9) holds. \square

Lemma 5.5. *Let $q \equiv 1 \pmod{3}$ be a prime power, and let m be the greatest divisor of $q^3 - 1$ co-prime to $q - 1$. Then*

$$\theta(m) > \frac{1}{q^{\frac{1}{6}}}.$$

Proof. Observe first that, if l is a prime divisor of m , then l is congruent to 1 modulo 6 and hence $l \geq 7$. Since $x - x^{\frac{1}{12}} - 1 > 0$ holds for $x \geq 7$, it follows that $\theta(p^k) = \theta(p) = \frac{p-1}{p} > \frac{1}{p^{\frac{1}{12}}} \geq \frac{1}{(p^k)^{\frac{1}{12}}}$ where $p \geq 7$ is prime and $k \in \mathbb{N}$. Thus by multiplicativity, $\theta(m) > \frac{1}{m^{\frac{1}{12}}}$. Since $3m \leq \frac{q^3-1}{q-1} < (q+1)^2$, it follows that $q > 3^{1/2}m^{1/2} - 1$, and so $q \geq m^{\frac{1}{2}}$ for all q . Hence $\frac{1}{q} \leq \frac{1}{m^{\frac{1}{2}}}$, and so $\theta(m) > \frac{1}{m^{\frac{1}{12}}} \geq \frac{1}{q^{\frac{1}{6}}}$. \square

Proposition 5.6. *Let $q \equiv 1 \pmod{3}$ be a prime power. Then $(q, 3)$ is a PFNT pair for all $q \geq 252,950$.*

Proof. By Lemma 5.4, $\pi(m, M) > 0$ if

$$(5.10) \quad \begin{aligned} & \pi(1, 1) \left(\theta(m) - \frac{2}{q} \right) \\ & > 3\theta(m)(W(m) - 1) \left(1 - \frac{1}{q} \right) q^{\frac{5}{2}} + 2q^{\frac{5}{2}} \left(1 - \frac{3}{q} - \frac{2}{q^2} \right) + 2q^2 \left(1 - \frac{3}{q} \right). \end{aligned}$$

Then by Lemma 3.1, $\pi(m, M) > 0$ if

(5.11)

$$\theta(m)\left(q^3-3W(m)\left(1-\frac{1}{q}\right)q^{\frac{5}{2}}-1\right)>2q^{\frac{5}{2}}\left(1-\frac{6}{q}+\frac{1}{q^2}\right)+2q^2\left(2-\frac{3}{q}\right)-\frac{2}{q}.$$

By Lemma 4.1, $W(m) \leq \frac{c_m q^{\frac{1}{2}}}{3^{\frac{1}{6}}(q-1)^{\frac{1}{6}}}$, where $c_m < 3.08$. Set $d := 3^{\frac{5}{6}}c_m$; then $3W(m) \leq \frac{dq^{\frac{1}{2}}}{(q-1)^{\frac{1}{6}}}$, and so $3W(m)(\frac{q-1}{q})q^{\frac{5}{2}} \leq d(q-1)^{\frac{5}{6}}q^2$. Using this result and Lemma 5.5, we see that $\pi(m, M) > 0$ certainly if

(5.12)

$$\frac{1}{q^{\frac{1}{6}}}\{q^3-d(q-1)^{\frac{5}{6}}q^2-1\}>2q^{\frac{5}{2}}\left(1-\frac{6}{q}+\frac{1}{q^2}\right)+2q^2\left(2-\frac{3}{q}\right)-\frac{2}{q},$$

i.e., if

(5.13)

$$q>d(q-1)^{\frac{5}{6}}+2q^{\frac{2}{3}}\left(1-\frac{6}{q}+\frac{1}{q^2}\right)+2q^{\frac{1}{6}}\left(2-\frac{3}{q}\right)+\frac{1}{q^2}.$$

Take $c_m = 3.08$ and set $d = 7.70$ in inequality (5.13). Then inequality (5.13) holds for all $q \geq 252,950$.

In order to establish the result for smaller prime powers q , we will use the following sufficient condition, which arises from the application of the sieve with atomic divisors.

Once again we shall adopt the convention that all unmarked summation signs have index i running from $i = 1$ to s .

Lemma 5.7. *The following is a sufficient condition for $(q, 3)$ to be a PFNT pair. When $q \equiv 1 \pmod{3}$,*

(5.14)

$$\sqrt{q}>\frac{(3s+2)-\frac{(3s+6)}{q}-\frac{4}{q^2}-3(1-\frac{1}{q})\sum\frac{1}{p_i}+\frac{2}{\sqrt{q}}(1-\frac{3}{q})}{1-\sum\frac{1}{p_i}-\frac{2}{q}}+3(1-\frac{1}{q})+\frac{1}{q^{\frac{5}{2}}}$$

(where $m = p_1^{\alpha_1} \dots p_s^{\alpha_s}$).

Proof Let $m = p_1^{\alpha_1} \dots p_s^{\alpha_s}$, where p_1, \dots, p_s are distinct primes and $s \in \mathbb{N}$ (recall that the values of the α_i will be irrelevant here). Apply the sieve in the form

(5.15)

$$\pi(m, M) \geq \pi(p_1, 1) + \dots + \pi(p_s, 1) + \pi(1, x - \gamma) + \pi(1, x - \gamma^2) - (s + 1)\pi(1, 1).$$

Using the results of Lemma 5.1 and Corollary 3.3, $\pi(m, M) > 0$ if

(5.16)

$$\begin{aligned} &\pi(1, 1)\left(1-\sum\frac{1}{p_i}-\frac{2}{q}\right) \\ &\quad -2q^{\frac{5}{2}}\left(1-\frac{3}{q}-\frac{2}{q^2}\right)-2q^2\left(1-\frac{3}{q}\right)-3q^{\frac{5}{2}}\left(1-\frac{1}{q}\right)\sum\left(1-\frac{1}{p_i}\right)>0, \end{aligned}$$

i.e., if

(5.17)

$$\pi(1, 1)>\frac{q^{\frac{5}{2}}((3s+2)-\frac{(3s+6)}{q}-\frac{4}{q^2}-3(1-\frac{1}{q})\sum\frac{1}{p_i})+2q^2(1-\frac{3}{q})}{1-\sum\frac{1}{p_i}-\frac{2}{q}},$$

and so, using Lemma 3.1, certainly if

$$(5.18) \quad q > \frac{\sqrt{q}((3s+2) - \frac{(3s+6)}{q} - \frac{4}{q^2}) - 3\sqrt{q}(1 - \frac{1}{q}) \sum \frac{1}{p_i} + 2(1 - \frac{3}{q})}{1 - \sum \frac{1}{p_i} - \frac{2}{q}} + 3\sqrt{q}(1 - \frac{1}{q}) + \frac{1}{q^2}.$$

Observe that the inequalities of Lemma 5.7 are of use only when the denominator $1 - \sum \frac{1}{p_i} - \frac{3}{q} > 0$; in particular, it is necessary to have $\sum \frac{1}{p_i} < 1$. However, since all prime powers q that are congruent to 1 modulo 3 and less than 252,950 have $s \leq 7$, and all prime divisors of m are congruent to 1 modulo 6, the denominator is always positive in this case. \square

Proposition 5.8. *Suppose $q \equiv 1 \pmod{3}$ and $q \leq 252,950$, but $q \notin \{7, 13, 16, 19, 25, 31, 37, 43, 49, 61, 64, 67, 79, 109, 121, 163, 211, 256\}$. Then $(q, 3)$ is a PFNT pair.*

Proof. For $q > 4$, observe that

$$\sum \frac{1}{p_i} \geq \frac{3}{q^2} - \frac{3}{q^3},$$

since $\sum \frac{1}{p_i} \geq \frac{3}{q^2+q+1} = \frac{3}{q^2}(1 - \frac{1}{q} + \frac{1}{q(q^2+q+1)})$. Using this lower bound in Lemma 5.7, the desired result holds if

$$(5.19) \quad \sqrt{q} > \frac{(3s+2) - \frac{(3s+6)}{q} - \frac{13}{q^2} + \frac{18}{q^3} + \frac{2}{\sqrt{q}}(1 - \frac{3}{q})}{1 - \sum \frac{1}{p_i} - \frac{2}{q}} + 3 \left(1 - \frac{1}{q}\right) + \frac{1}{q^{\frac{5}{2}}}.$$

An upper bound is required for $\sum \frac{1}{p_i}$, say $\sum \frac{1}{p_i} \leq K(q)$ for some function K . In general, to simplify calculations, the crude estimate

$$(5.20) \quad \sum_{i=1}^s \frac{1}{p_i} \leq \sum_{i=1}^s \frac{1}{p[i]}$$

will be used, where $p[i]$ is the i th prime congruent to 1 modulo 6, as in Section 4. (More precise values may be taken in specific cases.)

Observe that the desired result certainly holds when

$$(5.21) \quad \sqrt{q} > \frac{(3s+2) + \frac{2}{\sqrt{q}} + \frac{18}{q^3}}{1 - \sum \frac{1}{p_i} - \frac{2}{q}} + 3 + \frac{1}{q^{\frac{5}{2}}},$$

and, for fixed s , the function of q on the right side of (5.21) clearly decreases as q increases. Hence to prove for a given s that the result is true for $q \geq q_0$ (some $q_0 \in \mathbb{N}$), it is sufficient to show that inequality (5.21) holds for $q = q_0$.

For $q \leq 252,950$, we have $s \leq 7$. Using the basic estimate

$$(5.22) \quad \sum \frac{1}{p_i} \leq \frac{1}{7} + \frac{1}{13} + \frac{1}{19} + \frac{1}{31} + \frac{1}{37} + \frac{1}{43} + \frac{1}{61} < 0.3714,$$

inequality (5.21) holds with $s = 7$ for relevant $q > 1580$, hence for all $q \geq 1597$. Now, for prime powers $q \equiv 1 \pmod{3}$ less than 1580, it happens that $s \leq 4$; in fact, except for the two values $q = 919$ and $q = 1369$, $s \leq 3$. Using the estimate

$$(5.23) \quad \sum \frac{1}{p_i} \leq \frac{1}{7} + \frac{1}{13} + \frac{1}{19} + \frac{1}{31} < 0.3047,$$

inequality (5.21) holds with $s = 4$ for $q > 546$ and hence for all $q \geq 547$. For $s = 3$, use of inequality (5.20) in (5.21) establishes the result for $q > 339$, i.e., $q \geq 343$. For $q = 277$ ($m = 7 \cdot 19 \cdot 193$) and $q = 289$ ($m = 7 \cdot 13 \cdot 307$), use of exact

values in Lemma 5.7 establishes the result. However, this approach is insufficient for $\{121, 163, 211, 256\}$. In the $s = 2$ case, inequality (5.21) establishes the result for $q > 185$, i.e., $q \geq 193$, when applied with the approximation of (5.20), and for $q = 169$ ($m = 61 \cdot 157$) and $q = 181$ ($m = 79 \cdot 139$) when exact values are used in (5.21) (respectively, $181 > 153.49$ and $169 > 148.80$). Use of Lemma 5.7 suffices for $q = 139$ ($m = 13 \cdot 499$), since $139 > 137.14$. Outstanding exceptions in the $s = 2$ case are $\{16, 25, 37, 49, 61, 64, 67, 79, 109\}$. When $s = 1$, replacing p_1 by 7 in inequality (5.21) establishes the result for $q > 86$, i.e., $q \geq 97$; use of exact $p_1 (= m)$ deals with the case $q = 73$ ($m = 1801$). The remaining exceptions with $s = 1$ are $\{7, 13, 19, 31, 43\}$. \square

6. COMPUTATIONAL STRATEGY FOR REMAINING CASES

To deal with the 34 cases remaining after Propositions 4.5 and 5.8, we use the computer package MAPLE (version 6) to search the field E for m -free elements with norms and traces equal to the required values. (For reference, the set of exceptional q is as follows: $\{3, 5, 7, 8, 9, 11, 13, 16, 17, 19, 23, 25, 29, 31, 32, 37, 43, 47, 49, 53, 61, 64, 67, 79, 81, 107, 109, 121, 137, 149, 163, 191, 211, 256\}$.)

The following lemma allows us to simplify our computational strategy in some cases for which the PFNT problem reduces to the PNT.

Lemma 6.1. *Let q be a prime power, $q \not\equiv 1 \pmod{3}$. Denote by $Z_{\alpha,\beta}(m)$ the number of elements $w \in E$ that are m -free and have $\text{Tr}_{E/F}(w) = \alpha$, $N_{E/F}(w) = \beta$ ($\alpha, \beta \in F$). Suppose*

$$Z_{1,b}(m) > 0 \quad \forall b \in F^*.$$

Then $(q, 3)$ is a PNT pair.

Proof. To prove that $(q, 3)$ is a PNT pair, we must show that $N(m, 1) > 0$, i.e., that $Z_{a,b}(m) > 0$ for all $a, b \in F$, $a \neq 0$, b primitive. We prove the (stronger) result

$$Z_{a,b}(m) > 0 \quad \forall a, b \in F^*.$$

If $a = 1$, there is nothing to prove. Otherwise, set $b^* := \frac{b}{a^3} \in F^*$. Since $Z_{1,b^*}(m) > 0$, there exists an element $\zeta \in E$ such that ζ is m -free, $\text{Tr}_{E/F}(\zeta) = 1$, and $N_{E/F}(\zeta) = b^*$. Then $\alpha := a\zeta$ is also m -free, and has $\text{Tr}_{E/F}(\alpha) = a$ and $N_{E/F}(\alpha) = b$.

Use of Lemma 6.1 reduces the number of necessary tests from $(q - 1)\phi(q - 1)$ (testing each pair (a, b) , b primitive) to $q - 1$ (testing each pair $(1, b)$, b nonzero). Since the condition involved is stronger than the PNT condition, this simplification is only of practical use in those cases when $q - 1$ is prime, or $\phi(q - 1)$ is not too much smaller than $q - 1$. However, it is successful in dealing with all $q \not\equiv 1 \pmod{3}$ up to $q = 32$. For larger values of q , we must search E explicitly.

In the PNT case, the desired result holds without exception for all $q \not\equiv 1 \pmod{3}$ remaining from the previous sections.

As an illustration, we display the relevant cubic polynomials for the case when $q = 5$. The following table lists eight cubic polynomials over $F = \text{GF}(5)$ whose roots $\alpha \in E = \text{GF}(5^3)$ are primitive and free with norm and trace equal to b and a respectively.

(a, b)	Relevant PFNT cubic
(1,2)	$x^3 + 4x^2 + 3$
(1,3)	$x^3 + 4x^2 + x + 2$
(2,2)	$x^3 + 3x^2 + 2x + 3$
(2,3)	$x^3 + 3x^2 + 2$
(3,2)	$x^3 + 2x^2 + 3$
(3,3)	$x^3 + 2x^2 + 2x + 2$
(4,2)	$x^3 + x^2 + x + 3$
(4,3)	$x^3 + x^2 + 2$

The cubic polynomials given in the table for $(a, b) = (1, 2)$ and $(4, 3)$ are in fact unique. Thus, when $q = 5$ and $n = 3$, we observe that in some sense the PFNT property “only just” holds.

In the case when $q \equiv 1 \pmod{3}$, we search through E explicitly for elements with the required properties. The following lemma lets us reduce the number of pairs (a, b) that must be tested from $(q-1)\phi(q-1)$ to $\frac{1}{3}(q-1)\phi(q-1)$. \square

Lemma 6.2. *Let $q \equiv 1 \pmod{3}$, and set $k := \frac{q-1}{3}$. Suppose that there exist free, primitive $\alpha \in E$ such that $\text{Tr}_{E/F}(\alpha) = a$ and $N_{E/F}(\alpha) = b$, for all pairs (a, b) where b is a primitive element of F and $a \in \{1, \beta, \beta^2, \dots, \beta^{k-1} : \beta \text{ a fixed primitive element of } F\}$. Then there exist free, primitive $\alpha \in E$ such that $\text{Tr}_{E/F}(\alpha) = a$ and $N_{E/F}(\alpha) = b$, for all pairs (a, b) where a is a nonzero element of F and b is a primitive element of F .*

Proof. Fix a primitive element β of F . Observe that F^* may be partitioned into k cosets of the subgroup $H := \{1, \beta^k, \beta^{2k}\}$ of cube roots of unity, namely $H, \beta H, \dots, \beta^2 H, \dots, \beta^{k-1} H$. The result follows since $\text{Tr}_{E/F}(h\gamma) = h \text{Tr}_{E/F}(\gamma)$, $N(h\gamma) = h^3 N_{E/F}(\gamma)$ for all $\gamma \in E$, $h \in F$.

Without exception, for all $q \equiv 1 \pmod{3}$ remaining from the previous section, $(q, 3)$ is found to be a PFNT pair.

In closing we remark that, for each of the larger values of q amongst the set of exceptions, the computations to check all the possibilities took several hours to run. (In the case of $q = 256$, the original computation time of several weeks was reduced to a few hours by reprogramming.) This vindicates the efforts we have made to solve the problem theoretically in as many cases as possible. \square

REFERENCES

- [Cal] L. Carlitz, *Primitive roots in a finite field*, Trans. Amer. Math. Soc. **73** (1952), 373-382. MR **14**:539a
- [Ca2] L. Carlitz, *Some problems involving primitive roots in a finite field*, Proc. Nat. Acad. Sci. U.S.A. **38** (1952), 314-318, 618. MR **14**:250f
- [Co] S. D. Cohen, *Gauss sums and a sieve for generators of Galois fields*, Publ. Math. Debrecen **56** (2000), 293-312. MR **2001e**:11120
- [CoHa1] S. D. Cohen and D. Hachenberger, *Primitive normal bases with prescribed trace*, Appl. Algebra Engrg. Comm. Comp. **9** (1999), 383-403. MR **2000c**:11198
- [CoHa2] S. D. Cohen and D. Hachenberger, *Primitivity, freeness, norm and trace*, Discrete Math. **214** (2000), 135-144. MR **2000j**:11190
- [CoHu1] S. D. Cohen and S. Huczynska, *The primitive normal basis theorem — without a computer*, J. London Math. Soc. **67** (2003), 41-56.
- [CoHu2] S. D. Cohen and S. Huczynska, *Primitive free quartics with specified norm and trace*, Acta Arith. (to appear).

- [Da] H. Davenport, *Bases for finite fields*, J. London Math. Soc. **43** (1968), 21-49. MR **37**:2729
- [Ka] N. M. Katz, *Estimates for Soto-Andrade sums*, J. reine angew. Math. **438** (1993), 143-161. MR **94h**:11109
- [LeSc] H. W. Lenstra, Jr. and R. J. Schoof, *Primitive normal bases for finite fields*, Math. Comp. **48** (1987), 217-231. MR **88c**:11076
- [LiNi] R. Lidl and H. Niederreiter, *Finite Fields*, Addison-Wesley, Reading, Massachusetts, (1983); *2nd edition*: Cambridge University Press, Cambridge (1997). MR **97i**:11115

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GLASGOW, GLASGOW G12 8QW, SCOTLAND
Current address: School of Informatics, University of Edinburgh, Edinburgh EH8 9LE, Scot-

land

E-mail address: shuczyns@inf.ed.ac.uk

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GLASGOW, GLASGOW G12 8QW, SCOTLAND

E-mail address: sdc@maths.gla.ac.uk

D-LOG AND FORMAL FLOW FOR ANALYTIC ISOMORPHISMS OF N-SPACE

DAVID WRIGHT AND WENHUA ZHAO

ABSTRACT. Given a formal map $F = (F_1, \dots, F_n)$ of the form $z + \text{higher-order terms}$, we give tree expansion formulas and associated algorithms for the D-Log of F and the formal flow F_t . The coefficients that appear in these formulas can be viewed as certain generalizations of the Bernoulli numbers and the Bernoulli polynomials. Moreover, the coefficient polynomials in the formal flow formula coincide with the strict order polynomials in combinatorics for the partially ordered sets induced by trees. Applications of these formulas to the Jacobian Conjecture are discussed.

1. INTRODUCTION

This work began as an effort to link and extend the results of [W2] and [Z], placing them in a common framework. Both of these papers deal with the formal inverse F^{-1} of a system of power series $F = (F_1, \dots, F_n)$; both give formulas for F^{-1} in terms of F , the former being a tree formula, the latter an exponential formula. This quest has led to a host of interesting connections, algorithms, formulas, and relationships with combinatorics, Bernoulli numbers, and Bernoulli polynomials.

The former paper deals with tree formulas as they apply to formal inverse, a thread which is also the main thrust of [BCW], [W1], [W2], and [CMTWW]. It has combinatoric connections with generating functions and enumeration techniques for trees. The general goal of power series inversion (sometimes called “reversion”, perhaps to distinguish functional inverse from multiplicative inverse) is as follows. Let $F = (F_1, \dots, F_n)$ with $F_i \in \mathbb{C}[[z_1, \dots, z_n]]$ for each i and $F_i = z_i + \text{terms of degree } \geq 2$. One seeks formulas for the unique $G_1, \dots, G_n \in \mathbb{C}[[z_1, \dots, z_n]]$ for which $G_i(F) = z_i$, for $i = 1, \dots, n$. Perhaps the first of these was the Lagrange Inversion Formula (see [St2], Chapter 5), which dealt with the case $n = 1$, and which was generalized (under a certain restrictive hypothesis) to all n by I. J. Good [Go] in 1960. Good then uses his formula for problems of enumerating certain trees. In fact, Good’s formula had been discovered and published by Jacobi in 1830 [Ja]. Another paper which appeared in 1960 was that of G. N. Raney [R], who also related formal inverse to trees. Raney’s work was generalized in [CMTWW], which also utilized the work of Jacobi. A general inversion formula was given by Abhyankar and Gurjar in 1974 [A], and this is the source from which

Received by the editors June 5, 2002 and, in revised form, January 3, 2003.

2000 *Mathematics Subject Classification.* Primary 14R10, 11B68; Secondary 14R15, 05C05.

Key words and phrases. D-log, formal flow of automorphisms, rooted trees, order polynomials, Bernoulli numbers and polynomials.

the tree formula of [BCW] was derived, with the hope of applying it to the Jacobian Conjecture. Other treatments of the subject of inversion are [HS], [Ge], and [Jo].

The tree formula of [BCW] expresses the formal inverse F^{-1} as an infinite \mathbb{Q} -linear combination of certain power series $\mathcal{P}_T \in \mathbb{C}[[z]]$, which are constructed using finite rooted trees T . This construction will be reviewed in §2 and a new (and quick) proof of the inversion formula, using the tools developed in this paper, will be presented in §5 (Theorem 5.1).

Amongst the results of the latter paper is the realization of F by an expression $F = \exp(A) \cdot z$, where $A = A(z)$, called the *D-Log* of F , is a differential operator uniquely determined by F and yielding the formal inverse as $F^{-1} = \exp(-A) \cdot z$. Furthermore, the *formal flow* $F_t = \exp(tA) \cdot z$ encodes all powers $F^{[n]}$ with $n \in \mathbb{Z}$ of the formal map F . The D-Log and the formal flow will be reviewed in §3.

A primary goal was to show that the D-Log A can also be expressed as a \mathbb{Q} -linear combination of the power series \mathcal{P}_T . This goal was attained, yielding a tree formula for the D-Log. Moreover, we discovered that the rational coefficients ϕ_T of this expression can be generated by an elegant recurrence relationship and possess some intriguing combinatorial properties. For example, the Bernoulli numbers appear amongst these coefficients.

This situation is placed in a larger context which incorporates formal inverse by considering the formal flow $F_t = \exp(tA) \cdot z$, where t is an indeterminate. For $n \in \mathbb{N}$, setting $t = n$ gives the n -fold composition $F \circ \dots \circ F$, and setting $t = -n$ gives the n -fold composition $F^{-1} \circ \dots \circ F^{-1}$. The system F_t can be written as a $\mathbb{Q}[t]$ -linear combination of the power series \mathcal{P}_T , producing for each rooted tree T a polynomial $\psi_T(t)$ having ϕ_T as the coefficient of t . Among these polynomials are the binomial polynomials $\binom{t}{m}$, for all positive integers m . We give an algorithm for calculating $\psi_T(t)$ using the difference operator Δ . This formula is used to establish the relationship of certain $\psi_T(t)$'s with the Bernoulli polynomials $B_m(t)$ via an integration formula. It is shown that $\psi_T(t)$ also provides an interesting combinatorial connection: It coincides with the strict order polynomial $\bar{\Omega}(P, t)$ (see [St1], Chapter 4) for $P = T$, which, for $t = m \in \mathbb{Z}^+$, counts the number of strict order-preserving maps from any partially ordered set P to the totally ordered set with m elements.

Acknowledgments. We would like to thank Professor John Shareshian for informing us of Theorem 4.5, and also Professor Steve Krantz for a helpful conversation on flow of analytic maps.

2. TREE OPERATIONS

2.1. Notation. By a *rooted tree* we mean a finite 1-connected graph with one vertex designated as its *root*. The 1-connectivity provides the notion of *distance* between two vertices, which is defined as the number of edges in the unique geodesic connecting the two. The *height* of a tree is defined to be the maximum distance of any vertex from the root. In a rooted tree there are natural ancestral relations between vertices. We say a vertex w is a *child* of vertex v if the two are connected by an edge and w lies further from the root than v . In the same situation, we say v is the *parent* of w . Note that a vertex may have several children, but only one parent. The root is the only vertex with no parent. A vertex is called a *leaf* if it has no children. When we speak of isomorphisms between rooted trees, we will always mean root-preserving isomorphisms.

With these notions in mind, we establish the following notation.

- (1) We let \mathbb{T} be the set of isomorphism classes of all rooted trees and, for $m \geq 1$ an integer, we let \mathbb{T}_m be the set of isomorphism classes of all rooted trees with m vertices. The latter is a finite set.
- (2) For any rooted tree T , we set the following notation:
 - rt_T denotes the root vertex of T .
 - $E(T)$ denotes the set of edges of T .
 - $V(T)$ denotes the set of vertices of T .
 - $L(T)$ denotes the set of leaves of T .
 - $v(T)$ (resp. $l(T)$) denotes the number of the elements of $V(T)$ (resp. $L(T)$).
 - $h(T)$ denotes the height of T .
 - α_T denotes the number of the elements of the automorphism group $\text{Aut}(T)$.
 - For $v \in V(T)$ we denote by $\alpha_{T,v}$ the size of the stabilizer of v in $\text{Aut}(T)$. Similarly, for $e \in E(T)$, we write $\alpha_{T,e}$ for the size of the stabilizer of e in $\text{Aut}(T)$.
 - For $e \in E(T)$ we denote by v_e and v'_e the two (distinct) vertices that are connected by e , with v_e being the one closest to the root.
 - For $v \in V(T)$ we denote by v^+ the set of vertices that are children of v .
 - For $v \in V(T)$ we define the *height* of v to be the number of edges in the (unique) geodesic connecting v to rt_T . The *height* of T is defined to be the maximum of the heights of its vertices.
 - For $v_1, \dots, v_r \in V(T)$, we write $T \setminus \{v_1, \dots, v_r\}$ for the graph obtained by deleting each of these vertices and all edges adjacent to these vertices.
- (3) A *rooted subtree* of a rooted tree T is defined as a connected subgraph of T containing rt_T , with $\text{rt}_{T'} = \text{rt}_T$. In this case we write $T' \leq T$. If $T' \neq T$, we write $T' < T$. If $T' < T$, we write $T \setminus T'$ for the graph obtained by deleting all vertices of T' and all edges adjacent to its vertices.
- (4) For any $k \geq 1$, we denote by C_k the rooted tree of height $k - 1$ having k vertices, and by S_k the rooted tree of height 1 having k leaves. We also set $S_0 = \circ$, the rooted tree with one vertex. We refer to the trees C_k as *chains* and the S_k as *shrubs*.

2.2. Power Series Given by a Rooted Tree. Let $\mathbb{C}[[z_1, \dots, z_n]] = \mathbb{C}[[z]]$ denote the ring of formal power series in n variables z_1, \dots, z_n over the complex numbers¹ \mathbb{C} . For $i = 1, \dots, n$ we will write D_i for the differential operator $\frac{\partial}{\partial z_i}$. The operators D_1, \dots, D_n are commuting derivations acting on the ring $\mathbb{C}[[z]]$.

Given a vector of power series $F = (F_1, \dots, F_n) \in \mathbb{C}[[z]]^n$, we write $F_i = z_i + H_i$ for $i = 1, \dots, n$, or just $F = z + H$.² In most applications the power series $H = (H_1, \dots, H_n)$ will involve only monomials of total degree 2 and higher, and we will often take H to be homogeneous of degree $d \geq 2$. However, these assumptions are

¹In this paper \mathbb{C} can always be replaced by any \mathbb{Q} -algebra.

²We should here acknowledge that in almost every other treatment of this subject the system F is written as $z - H$, which yields nicer looking formulas for the formal inverse of F . The reason for our choice is that the formulas involving the D-Log and formal flow, which will be developed in §3, come out better when we write $F = z + H$.

not necessary for what follows here. We will associate to each rooted tree a power series in n variables based on F (equivalently, on H).

For $T \in \mathbb{T}$, a *labeling* of T in the set $\{1, \dots, n\}$ is a function $f : V(T) \rightarrow \{1, \dots, n\}$. A rooted tree T with a labeling f is called a *labeled rooted tree*, denoted (T, f) . Given such, and given $F = z + H$ as above, we make the following definitions, for $v \in V(T)$:

- (1) $H_v = H_{f(v)}$;
- (2) $D_v = D_{f(v)}$;
- (3) $D_{v^+} = \prod_{w \in v^+} D_w$;
- (4) $P_{T,f} = \prod_{v \in V(T)} D_{v^+} H_v$.

Finally, we define systems of power series $P_T = (P_{T,1}, \dots, P_{T,n})$ and $\mathcal{P}_T = (\mathcal{P}_{T,1}, \dots, \mathcal{P}_{T,n})$ by summing over all labelings of T having a fixed label for the root

$$P_{T,i} = \sum_{\substack{f: V(T) \rightarrow \{1, \dots, n\} \\ f(\text{rt}_T) = i}} P_{T,f},$$

$$\mathcal{P}_{T,i} = \frac{1}{\alpha_T} P_{T,i}$$

for $i = 1, \dots, n$.

One notes that the systems of power series P_T and \mathcal{P}_T are dependent on the integer n and the system $H = (H_1, \dots, H_n) \in \mathbb{C}[[z]]^n$. They can be viewed as objects which determine functions $\mathbb{C}[[z_1, \dots, z_n]]^n \rightarrow \mathbb{C}[[z_1, \dots, z_n]]^n$ for all $n \geq 1$. We will write $P_T(H)$ and $\mathcal{P}_T(H)$ when we need to emphasize this dependence, or when we are dealing with more than one system H .

2.3. Stable Linear Independence. We begin by establishing an important independence property of the objects $\{P_T \mid T \in \mathbb{T}\}$.

Definition 2.1. We say that rooted trees T_1, \dots, T_k are stably linearly dependent if there exist $c_1, \dots, c_k \in \mathbb{C}$ such that $\sum_{i=1}^k c_i P_{T_i} = 0$ for any integer $n \geq 1$ and any homogeneous polynomial system $H = (H_1, \dots, H_n)$ in n variables. Otherwise, we say that T_i are stably linearly independent.

Remark 2.2. If H is homogeneous of degree d and if $T \in \mathbb{T}_m$, then $P_T(H)$ is homogeneous of degree $(d-1)m+1$. Thus if we partition $\{T_1, \dots, T_k\}$ according to the number of vertices in a tree, then T_1, \dots, T_k are stably linearly independent if and only if each partition is a stably linearly independent set of trees.

Lemma 2.3. Suppose that $\sum_{i=1}^k c_i P_{T_i}(H) = 0$ for any integer $n \geq 1$ and any homogeneous polynomial system H in n variables. Then $\sum_{i=1}^k c_i P_{T_i}(H) = 0$ for any system of power series $H = (H_1, \dots, H_n)$ in n variables.

Proof. We first prove it for any polynomial H (not necessarily homogeneous) in n variables by introducing a new variable z_{n+1} and homogenizing H using z_{n+1} . Call the resulting homogeneous system \tilde{H} . Setting $\tilde{H} = (\tilde{H}, H_{n+1} = 0)$, we have

$$\sum_{i=1}^k c_i P_{T_i}(H, z) = \sum_{i=1}^k c_i P_{T_i}(\tilde{H}, z)|_{z_{n+1}=1} = 0,$$

which proves the lemma for H a polynomial system. For an arbitrary system of power series H we note that if T is a tree with r edges, the homogeneous summands

of degree $\leq d$ in $P_T(H)$ depend only on the homogeneous summands of H having degree $\leq d+r$. Taking r to be the maximum of the numbers of edges in T_1, \dots, T_k , then all terms of degree $\leq d$ in $\sum_{i=1}^k c_i P_{T_i}(H)$ depend only the homogeneous summands of H having degree $\leq d+r$. Taking \widehat{H} to be the polynomial truncation of H of degree $d+r$, we see that $\sum_{i=1}^k c_i P_{T_i}(H)$ and $\sum_{i=1}^k c_i P_{T_i}(\widehat{H})$ coincide up through degree d . Since the latter is zero (\widehat{H} being a polynomial system) and d is arbitrary, we must have $\sum_{i=1}^k c_i P_{T_i}(H) = 0$. \square

Theorem 2.4. *Any rooted trees T_i ($i = 1, 2, \dots, k$) with $T_i \not\cong T_j$ for any $i \neq j$ are stably linearly independent.*

Before giving the proof we will define a polynomial system depending on a rooted tree. Given a rooted tree T with m vertices, we create variables z_1, \dots, z_m . Label the edges e_2, \dots, e_m and assign each variable z_i with $2 \leq i \leq m$ to the edge e_i . Label the vertices as follows: $v_1 = \text{rt}_T$, and for $i = 2, \dots, m$ let v_i be the vertex of e_i that is farthest from the root. For each vertex $v_i \in V(T)$, we define H_i to be the product of all the variables assigned to the edges connecting v_i with its children. (Thus if v_i is a leaf, we have $H_i = 1$.) Set $H_T = (H_1, \dots, H_m)$. We have

Lemma 2.5. *Let T and T' be two rooted trees with the same number of vertices. Then*

$$P_{T'}(H_T) = \begin{cases} (0, \dots, 0) & \text{if } T \not\cong T', \\ (\alpha_T, 0, \dots, 0) & \text{if } T \cong T'. \end{cases}$$

Proof. The following facts are not difficult to verify and provide a sketch of the proof: Each coordinate $H_{T,i}$ of H_T is a monomial that is linear or constant with respect to each variable z_i . Each coordinate is constant with respect to z_1 . Each variable z_i with $i \geq 2$ appears in precisely one coordinate $H_{T,j}$, and $i \neq j$. $P_{T'}(H_T)$ is a homogeneous system of degree zero, and must be equal to either 0 or 1. If a labeling $f : V(T') \rightarrow \{1, \dots, m\}$ is not bijective, then $P_{T',f} = 0$ since it would entail differentiating two different coordinates $H_{T,i}$ and $H_{T,j}$ with respect to the same variable, or differentiating some $H_{T,i}$ twice by the same variable, or differentiating some $H_{T,i}$ by z_i , all of which give zero. Moreover, if $f(\text{rt}_{T'}) \neq 1$, then $P_{T',f} = 0$, since it would entail differentiation by z_1 , and therefore $P_{T'}(H_T)$ is zero except possibly in the first coordinate.

With this it is not hard to show that, if $f : V(T') \rightarrow \{1, \dots, m\}$ is a labeling for which $P_{T',f} \neq 0$, then the function $V(T') \rightarrow V(T)$ defined by $w \mapsto v_{f(w)}$ gives an isomorphism of $\varphi : T' \rightarrow T$. Finally, the group $\text{Aut } T$ acts freely and transitively on the set of labelings $f : V(T) \rightarrow \{1, \dots, m\}$ for which $P_{T,f} \neq 0$. The lemma follows easily from these statements. \square

Proof of Theorem 2.4. Suppose that $\sum_{i=1}^k c_i P_{T_i}(z) = 0$ with $c_1 \neq 0$. Choose $H = H_{T_1}$. Then there must exist $j \neq 1$ such that $P_{T_j}(H_{T_1}) \neq 0$. By the lemma above, we have $T_1 \cong T_j$. \square

If $H = (H_1, \dots, H_n)$ is a system of power series such that each H_i has only terms of degree d and higher, the power series \mathcal{P}_T has only terms of degree $(d-1)v(T)+1$ and higher. Hence if $d \geq 2$, a sum of the form $\sum_{T \in \mathbb{T}} c_T \mathcal{P}_T$ makes sense, since only finitely many terms contribute to any specified homogeneous summand. With this observation, we state the following consequence of stable linear independence.

Corollary 2.6. *Suppose we have a collection $\{c_T\} \subset \mathbb{C}$ indexed by the rooted trees $T \in \mathbb{T}$ such that $\sum_{T \in \mathbb{T}} c_T \mathcal{P}_T = 0$ for any integer $n \geq 1$ and any system of power series $H = (H_1, \dots, H_n)$ with H having only terms of degree ≥ 2 . Then $c_T = 0$ for all $T \in \mathbb{T}$.*

Proof. We consider systems H that are homogeneous polynomial systems of degree $d \geq 2$. In this case \mathcal{P}_T is homogeneous of degree $(d-1)v(T)+1$. So the homogeneous summands of $\sum_{T \in \mathbb{T}} c_T \mathcal{P}_T$ are the finite sums $\sum_{T \in \mathbb{T}_N} c_T \mathcal{P}_T$ for $N \in \mathbb{N}$; so these must be zero. By Theorem 2.4 applied to the finite set of trees \mathbb{T}_N , we must have $c_T = 0$ for all $T \in \mathbb{T}_N$. □

Recall that we are writing D_i for the operator $\frac{\partial}{\partial z_i}$. We will denote by D the column vector $(D_1, \dots, D_n)^t$. We now define a differential operator on $\mathbb{C}[[z]]$ for each $T \in \mathbb{T}$.

Definition 2.7. For $T \in \mathbb{T}$, we denote by D_T the differential operator $P_T D = \sum_{i=1}^n P_{T,i} D_i$. We will write \mathcal{D}_T for the operator $\mathcal{P}_T D = \frac{1}{\alpha_T} D_T$.

2.4. Tree Surgery. We will now discuss some “surgical” procedures on trees. Given $T \in \mathbb{T}$ and $e \in E(T)$, the removal of the edge e from T gives a disconnected graph with two connected components which are trees. We denote by T_e the component containing rt_T , and by T'_e the other component. We give T_e and T'_e the structure of rooted trees by setting $\text{rt}_{T_e} = \text{rt}_T$ and $\text{rt}_{T'_e} = v'_e$.

Given rooted trees T and T' and $v \in V(T)$, we denote by

$$T' \text{--} \circ_v T$$

the tree obtained by connecting $\text{rt}_{T'}$ and v by a newly created edge, and setting $\text{rt}_{(T' \text{--} \circ_v T)} = \text{rt}_T$. We will refer to the newly created edge as the *connection edge* of $T' \text{--} \circ_v T$. Note that for any tree T and edge $e \in E(T)$ we have an obvious isomorphism $T \cong (T'_e \text{--} \circ_{v_e} T_e)$ which is the identity on T_e and T'_e .

Given $e, f \in E(T)$, we say “ f lies below e ”, and write $e \succ f$, if $f \in E(T_e)$. This merely says that f remains when we “strip away” e and T'_e . One can easily see that this relation is not transitive. However, if we write

$$e_1 \succ \cdots \succ e_r,$$

for $e_1, \dots, e_r \in E(T)$, we will mean by this that $e_i \succ e_j$ if $i < j$.

A sequence $\vec{e} = (e_1, \dots, e_r) \in E(T)^r$ with $e_1 \succ \cdots \succ e_r$ determines a sequence of subtrees $T_{\vec{e},1}, \dots, T_{\vec{e},r+1}$ as follows: Set $T_{\vec{e},1} = T'_{e_1}$ and let $S_2 = T_{e_1}$, noting that $e_2, \dots, e_r \in E(S_2)$. For $i = 1, \dots, r$, assume that $T_{\vec{e},1}, \dots, T_{\vec{e},i-1}, S_i$ are defined with $e_i, \dots, e_r \in E(S_i)$. Set $T_{\vec{e},i} = (S_i)'_{e_i}$ and $S_{i+1} = (S_i)_{e_i}$. Finally, set $T_{\vec{e},r+1} = S_{r+1}$.

For any integer $r \geq 1$ and $T \in \mathbb{T}$, create an indeterminate $Y_T^{(r)}$. Denote this set of variables (for all T and r) by Y . Extend the action of the operators D_T and \mathcal{D}_T to $\mathbb{C}[[z]][Y]$ by making each indeterminate of Y a constant.

Lemma 2.8. *Let $r, m \geq 1$ be integers and $S \in \mathbb{T}$. Then*

(2.1)

$$\begin{aligned} &\sum_{\substack{(T_1, \dots, T_r) \in \mathbb{T}^r \\ v(T_1) + \cdots + v(T_r) + v(S) = m}} \left[Y_{T_1}^{(1)} \mathcal{D}_{T_1} \right] \cdots \left[Y_{T_r}^{(r)} \mathcal{D}_{T_r} \right] \mathcal{P}_S \\ &= \sum_{T \in \mathbb{T}_m} \sum_{\substack{\vec{e} = (e_1, \dots, e_r) \in E(T)^r \\ e_1 \succ \cdots \succ e_r \\ T_{\vec{e},r+1} \cong S}} Y_{T_{\vec{e},1}}^{(1)} \cdots Y_{T_{\vec{e},r}}^{(r)} \mathcal{P}_T. \end{aligned}$$

Proof. Note that both sums are finite. So the expression makes sense for any $H \in \mathbb{C}[[z]]^n$.

We first consider the case $r = 1$. For $T' \in \mathbb{T}$ we have

$$D_{T'} P_S = \sum_{v \in V(S)} P_{(T' \multimap_v S)}.$$

Hence

$$\begin{aligned} \sum_{\substack{T' \in \mathbb{T} \\ v(T') + v(S) = m}} Y_{T'}^{(1)} D_{T'} P_S &= \sum_{\substack{T' \in \mathbb{T} \\ v(T') + v(S) = m}} \sum_{v \in V(S)} Y_{T'}^{(1)} P_{(T' \multimap_v S)} \\ &= \sum_{T \in \mathbb{T}_m} \sum_{T' \in \mathbb{T}} \sum_{\substack{v \in V(S) \\ (T' \multimap_v S) \cong T}} Y_{T'}^{(1)} P_{(T' \multimap_v S)}. \end{aligned}$$

For a fixed $T \in \mathbb{T}_m$ we wish to count the occurrences of P_T in the last expression. Toward this end, for $T' \in \mathbb{T}$, let

$$\begin{aligned} I_{T,T',S} &= \{v \in V(S) \mid (T' \multimap_v S) \cong T\}, \\ J_{T,T',S} &= \{\bar{e} \in E(T)/\text{Aut}(T) \mid T'_e \cong T', T_e \cong S \text{ (for any } e \text{ in } \bar{e})\}. \end{aligned}$$

We will define a function $\Phi : I_{T,T',S} \rightarrow J_{T,T',S}$ as follows: Given $v \in I_{T,T',S}$, choose an isomorphism $\varphi : (T' \multimap_v S) \xrightarrow{\cong} T$, and let e be the image under φ of the connection edge in $T' \multimap_v S$. Letting \bar{e} be the class of e in $E(T)/\text{Aut}(T)$, we clearly have $\bar{e} \in J_{T,T',S}$. To see that \bar{e} is independent of the choice of φ , suppose $\gamma : (T' \multimap_v S) \xrightarrow{\cong} T$ sends the connection edge to $f \in E(T)$. Then $\gamma\varphi^{-1}(e) = f$, hence $\bar{f} = \bar{e}$ in $E(T)/\text{Aut}(T)$. Therefore, we have a well-defined function Φ , which is obviously surjective.

We claim that for $v \in I_{T,T',S}$ the orbit of v under $\text{Aut}(S)$ is precisely the fiber of v under Φ . It is clear that if $w \sim v$ by the action of $\text{Aut}(S)$, then $(T' \multimap_w S) \cong (T' \multimap_v S) \cong T$, with the first isomorphism taking one connection edge to the other, which shows that $w \in I_{T,T',S}$. Choosing appropriate isomorphisms $(T' \multimap_w S) \xrightarrow{\rho} (T' \multimap_v S) \xrightarrow{\varphi} T$, we see that the image e of the connection edge of $T' \multimap_v S$ under φ is also the image of the connection edge of $T' \multimap_w S$ under $\varphi\rho$, hence $\Phi(w) = \Phi(v)$. Moreover, if $w \in I_{T,T',S}$ is any element for which $\Phi(w) = \Phi(v)$, then we have isomorphisms

$$(T' \multimap_w S) \xrightarrow{\gamma} T \xleftarrow{\varphi} (T' \multimap_v S)$$

such that the same $e \in E(T)$ is the image of both connection edges. (This can be achieved after modifying by an automorphism of T .) It follows that $\gamma^{-1}\varphi : (T' \multimap_v S) \rightarrow (T' \multimap_w S)$ carries one connection edge to the other; so it restricts to an automorphism of S sending v to w . Hence $w \sim v$. Therefore, the above sum can be written as

$$\begin{aligned} \sum_{T \in \mathbb{T}_m} \sum_{T' \in \mathbb{T}} \sum_{v \in I_{T,T',S}} Y_{T'}^{(1)} P_T \\ = \sum_{T \in \mathbb{T}_m} \sum_{T' \in \mathbb{T}} \sum_{\bar{e} \in J_{T,T',S}} s_{T_e}(v_e) Y_{T'}^{(1)} P_T \end{aligned}$$

$$= \sum_{T \in \mathbb{T}_m} \sum_{\substack{\bar{e} \in E(T)/\text{Aut}(T) \\ T_e \cong S}} s_{T_e}(v_e) Y_{T'_e}^{(1)} P_T$$

where $s_{T_e}(v_e)$ is the orbit size of v_e under the action of $\text{Aut } T_e$, for some (any) $e \in E(T)$ representing \bar{e} . The number of edges representing \bar{e} is $\alpha_T/\alpha_{T,e}$. Hence the inner sum can be altered to run over all $e \in E(T)$ at the cost of dividing by $\alpha_T/\alpha_{T,e}$, yielding

$$\sum_{T \in \mathbb{T}_m} \frac{1}{\alpha_T} \sum_{\substack{e \in E(T) \\ T_e \cong S}} \alpha_{T,e} s_{T_e}(v_e) Y_{T'_e}^{(1)} P_T.$$

An automorphism of T fixing $e \in E(T)$ restricts to an automorphism of T'_e and an automorphism of T_e fixing v_e . Conversely, given the latter pair we get a unique automorphism of T preserving e . It follows that $\alpha_{T,e} = \alpha_{T'_e} \alpha_{T_e, v_e}$. Also we have $s_{T_e, v_e} = \alpha_{T'_e}/\alpha_{T_e, v_e}$. Incorporating these facts and putting together the above equalities, we get

$$\sum_{\substack{T' \in \mathbb{T} \\ v(T') + v(S) = m}} Y_{T'}^{(1)} D_{T'} P_S = \sum_{T \in \mathbb{T}_m} \frac{1}{\alpha_T} \sum_{\substack{e \in E(T) \\ T_e \cong S}} \alpha_{T'_e} \alpha_S Y_{T'_e}^{(1)} P_T.$$

Dividing the equation by α_S and substituting $\frac{1}{\alpha_R} Y_R^{(1)}$ for $Y_R^{(1)}$ for each $R \in \mathbb{T}$ yields

$$\sum_{\substack{T' \in \mathbb{T} \\ v(T') + v(S) = m}} Y_{T'}^{(1)} D_{T'} P_S = \sum_{T \in \mathbb{T}_m} \sum_{\substack{e \in E(T) \\ T_e \cong S}} Y_{T'_e}^{(1)} P_T,$$

which is precisely the assertion of the lemma for $r = 1$.

For $r \geq 2$ we apply induction as follows:

$$\begin{aligned} & \sum_{\substack{(T_1, \dots, T_r) \in \mathbb{T}^r \\ v(T_1) + \dots + v(T_r) + v(S) = m}} \left[Y_{T_1}^{(1)} D_{T_1} \right] \cdots \left[Y_{T_r}^{(r)} D_{T_r} \right] P_S \\ &= \sum_{T_1 \in \mathbb{T}} Y_{T_1}^{(1)} D_{T_1} \sum_{\substack{(T_2, \dots, T_r) \in \mathbb{T}^{r-1} \\ v(T_2) + \dots + v(T_r) + v(S) = m - v(T_1)}} \left[Y_{T_2}^{(2)} D_{T_2} \right] \cdots \left[Y_{T_r}^{(r)} D_{T_r} \right] P_S. \end{aligned}$$

Applying induction and a substitution of variables $Y_t^{(i+1)}$ for $Y_t^{(i)}$ to the inner sum, this equals

$$\begin{aligned} & \sum_{T_1 \in \mathbb{T}} Y_{T_1}^{(1)} D_{T_1} \sum_{R \in \mathbb{T}_{m-v(T_1)}} \sum_{\substack{\bar{e} = (e_1, \dots, e_{r-1}) \in E(R)^{r-1} \\ e_1 \succ \dots \succ e_{r-1} \\ T_{\bar{e}, r} \cong S}} Y_{T_{\bar{e}, 1}}^{(2)} \cdots Y_{T_{\bar{e}, r-1}}^{(r)} P_R \\ &= \sum_{\substack{T_1, R \in \mathbb{T} \\ v(T_1) + v(R) = m}} Y_{T_1}^{(1)} D_{T_1} \sum_{\substack{\bar{e} = (e_1, \dots, e_{r-1}) \in E(R)^{r-1} \\ e_1 \succ \dots \succ e_{r-1} \\ T_{\bar{e}, r} \cong S}} Y_{T_{\bar{e}, 1}}^{(2)} \cdots Y_{T_{\bar{e}, r-1}}^{(r)} P_R \\ &= \sum_{R \in \mathbb{T}} \left[\sum_{\substack{T_1 \in \mathbb{T} \\ v(T_1) + v(R) = m}} Y_{T_1}^{(1)} D_{T_1} P_R \right] \sum_{\substack{\bar{e} = (e_1, \dots, e_{r-1}) \in E(R)^{r-1} \\ e_1 \succ \dots \succ e_{r-1} \\ T_{\bar{e}, r} \cong S}} Y_{T_{\bar{e}, 1}}^{(2)} \cdots Y_{T_{\bar{e}, r-1}}^{(r)}. \end{aligned}$$

Now we apply the case $r = 1$ to the bracketed expression to obtain

$$\begin{aligned} & \sum_{R \in \mathbb{T}} \left[\sum_{T \in \mathbb{T}_m} \sum_{\substack{e \in E(T) \\ T_e \cong R}} Y_{T'_e}^{(1)} \mathcal{P}_T \right] \sum_{\substack{\vec{e} = (e_1, \dots, e_{r-1}) \in E(R)^{r-1} \\ e_1 \succ \dots \succ e_{r-1} \\ T_{\vec{e}, r} \cong S}} Y_{T_{\vec{e}, 1}}^{(2)} \dots Y_{T_{\vec{e}, r-1}}^{(r)} \\ & \cdot \sum_{T \in \mathbb{T}_m} \sum_{R \in \mathbb{T}} \sum_{\substack{e \in E(T) \\ T_e \cong R}} \sum_{\substack{\vec{e} = (e_1, \dots, e_{r-1}) \in E(R)^{r-1} \\ e_1 \succ \dots \succ e_{r-1} \\ T_{\vec{e}, r} \cong S}} Y_{T'_e}^{(1)} Y_{T_{\vec{e}, 1}}^{(2)} \dots Y_{T_{\vec{e}, r-1}}^{(r)} \mathcal{P}_T \\ & = \sum_{T \in \mathbb{T}_m} \sum_{\substack{\vec{e} = (e_1, \dots, e_r) \in E(T)^r \\ e_1 \succ \dots \succ e_r \\ T_{\vec{e}, r+1} \cong S}} Y_{T_{\vec{e}, 1}}^{(1)} \dots Y_{T_{\vec{e}, r}}^{(r)} \mathcal{P}_T \end{aligned}$$

which completes the proof. \square

Suppose the system of power series $H = (H_1, \dots, H_n)$ has the property that each H_i involves only monomials of degree ≥ 2 in z_1, \dots, z_n . Then one easily verifies that for $T \in \mathbb{T}$, \mathcal{P}_T involves only monomials of degree $\geq v(T) + 1$. It follows that for a monomial M in z of degree m , $\mathcal{D}_T \cdot M$ involves only monomials of degree $\geq m + v(T)$. Therefore, infinite sums such as $\sum_{T \in \mathbb{T}} \mathcal{P}_T$ and $\sum_{T \in \mathbb{T}} \mathcal{D}_T$ make sense in this situation. The following two corollaries of Lemma 2.8 are based on this observation. The equations in both corollaries take place in the ring $\mathbb{C}[Y][[z]]$, where Y represents the infinite set of variables $\{Y_T^{(i)} \mid T \in \mathbb{T}, i \in \mathbb{Z}^+\}$.

Corollary 2.9. *Suppose that the system of power series H involves only monomials of degree ≥ 2 . Let $r \geq 1$ be an integer and $S \in \mathbb{T}$. Then*

$$\begin{aligned} & \sum_{(T_1, \dots, T_r) \in \mathbb{T}^r} \left[Y_{T_1}^{(1)} \mathcal{D}_{T_1} \right] \dots \left[Y_{T_r}^{(r)} \mathcal{D}_{T_r} \right] \mathcal{P}_S \\ (2.2) \quad & = \sum_{T \in \mathbb{T}} \sum_{\substack{\vec{e} = (e_1, \dots, e_r) \in E(T)^r \\ e_1 \succ \dots \succ e_r \\ T_{\vec{e}, r+1} = S}} Y_{T_{\vec{e}, 1}}^{(1)} \dots Y_{T_{\vec{e}, r}}^{(r)} \mathcal{P}_T. \end{aligned}$$

Proof. We simply sum (2.1) over all $m \geq 1$, noting the convergence of the sums by the observations above. \square

Corollary 2.10. *Suppose the system of power series H involves only monomials of degree ≥ 2 . Let $k \geq 2$ be an integer. Then*

$$\begin{aligned} & \sum_{(T_1, \dots, T_k) \in \mathbb{T}^k} \left[Y_{T_1}^{(1)} \mathcal{D}_{T_1} \right] \dots \left[Y_{T_{k-1}}^{(k-1)} \mathcal{D}_{T_{k-1}} \right] \left[Y_{T_k}^{(k)} \mathcal{P}_{T_k} \right] \\ (2.3) \quad & = \sum_{\substack{T \in \mathbb{T} \\ v(T) \geq 2}} \sum_{\substack{\vec{e} = (e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} Y_{T_{\vec{e}, 1}}^{(1)} \dots Y_{T_{\vec{e}, k}}^{(k)} \mathcal{P}_T. \end{aligned}$$

Proof. We apply Corollary 2.9, multiplying both sides of (2.2) by $Y_S^{(r+1)}$, setting $k = r + 1$, summing over all $S \in \mathbb{T}$. Note that the singleton tree contributes 0 in (2.2) for any $r \geq 1$, and thus the qualifier $v(T) \geq 2$ in (2.3). \square

3. D-LOG AND FORMAL FLOW

We will henceforth be restricting our attention to systems of power series $F = (F_1, \dots, F_n) \in \mathbb{C}[[z]]^n$ of the form $F_i = z_i + H_i$ with H_i involving only monomials of degree 2 and higher, for $i = 1, \dots, n$. We refer to this condition by saying “ F is of the form *identity plus higher*.” Such a system determines a \mathbb{C} -algebra automorphism of $\mathbb{C}[[z]]$, namely the automorphism that sends z_i to F_i for $i = 1, \dots, n$.

3.1. The D-Log. The following proposition appears as Proposition 2.1 in [Z].

Proposition 3.1. *For any $F = (F_1, F_2, \dots, F_n) \in \mathbb{C}[[z]]^n$ of the form identity plus higher, there exists a unique system of power series*

$$a = (a_1, a_2, \dots, a_n) \in \mathbb{C}[[z]]^n$$

involving only monomials of degree 2 and higher such that, letting $A = aD = \sum_{i=1}^n a_i D_i$, we have

(3.1)
$$\exp(A) \cdot z = F$$

where

$$\exp(A) = \sum_{k=0}^\infty \frac{A^k}{k!}$$

and $z = (z_1, \dots, z_n)$.

The reader will easily verify that the infinite sum $\exp(A) \cdot Q$ makes sense for any $Q \in \mathbb{C}[[z]]^n$ due to the fact that, for any integer $d \geq 0$, only finitely many terms $\frac{A^k}{k!} \cdot Q$ contribute to the degree d homogeneous summand. This is due to the fact that a involves only terms of degree 2 and higher.

Remark 3.2. It is well known that the exponential of a derivation on any \mathbb{Q} -algebra, when it makes sense, is an automorphism of that algebra. Any subring lying in the kernel of the derivation will be fixed by this automorphism. It follows from this fact, the comment above, and Proposition 3.1 that $\exp(A)$ is the \mathbb{C} -algebra automorphism of $\mathbb{C}[[z]]$ that sends z_i to F_i , for $i = 1, \dots, n$.

Definition 3.3. We call the unique system of power series $a = (a_1, a_2, \dots, a_n)$ obtained above the *Differential Log* or *D-Log* of the formal system F .

3.2. Coefficients ϕ_T of the D-Log.

Theorem 3.4. *There exists a unique set of rational numbers $\{\phi_T\}$ indexed by the set of rooted trees $T \in \mathbb{T}$, such that*

(3.2)
$$a = \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T.$$

These numbers satisfy, and are uniquely determined by, the following properties:

(3.3)
$$\begin{aligned} &\phi_T = 1 \text{ when } v(T) = 1 \text{ (i.e., } T = \circ, \text{ the singleton tree),} \\ &\phi_T = - \sum_{k=2}^{v(T)} \frac{1}{k!} \sum_{\substack{\vec{e} = (e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \phi_{T_{\vec{e},2}} \cdots \phi_{T_{\vec{e},k}} \text{ when } v(T) \geq 2. \end{aligned}$$

The latter formula can be restated as

$$(3.4) \quad \sum_{k=1}^{v(T)} \frac{1}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \phi_{T_{\vec{e},2}} \cdots \phi_{T_{\vec{e},k}} = 0.$$

(Here we must interpret the $k = 1$ summand as ϕ_T .)

Proof. Let us define ϕ_T by (3.3) and set $a' = \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T$, $A' = a'D$. Then $A' = \sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T$. We have

$$\begin{aligned} \exp(A') \cdot z &= \sum_{k=0}^{\infty} \frac{A'^k}{k!} \cdot z \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(\sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T \right)^k \cdot z \\ &= z + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}^k} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_k} \mathcal{D}_{T_k}] \cdot z \\ &= z + \sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T \cdot z + \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}^k} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_k} \mathcal{D}_{T_k}] \cdot z. \end{aligned}$$

Using the fact that $\mathcal{D}_T \cdot z = \mathcal{P}_T$:

$$= z + \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T + \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}^k} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_{k-1}} \mathcal{D}_{T_{k-1}}] [\phi_{T_k} \mathcal{P}_{T_k}].$$

Applying Corollary 2.10, substituting $Y_{T_i}^{(i)} = \phi_{T_i}$:

$$= z + \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T + \sum_{k=2}^{\infty} \frac{1}{k!} \sum_{\substack{T \in \mathbb{T} \\ v(T) \geq 2}} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \cdots \phi_{T_{\vec{e},k}} \mathcal{P}_T.$$

Letting S be the singleton tree:

$$= z + \phi_S \mathcal{P}_S + \sum_{\substack{T \in \mathbb{T} \\ v(T) \geq 2}} \left(\sum_{k=1}^{v(T)} \frac{1}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \cdots \phi_{T_{\vec{e},k}} \right) \mathcal{P}_T.$$

Since $\mathcal{P}_S = H$, and, by definition, $\phi_S = 1$ and the sum in parentheses is 0:

$$= z + H = F.$$

By the uniqueness property of a we must have $a' = a$. The uniqueness of the expression (3.2) for a follows from Theorem 2.4. \square

Chains and Shrubs. Two special types of trees are the “chains” and the “shrubs”, mentioned in §2. Given an integer $n \geq 1$ we let $C_n \in \mathbb{T}_n$ be the *chain* with n vertices, which is the unique rooted tree in \mathbb{T}_n of height $n - 1$. For $n \geq 0$ we let $S_n \in \mathbb{T}_{n+1}$ be the *shrub* with $n + 1$ vertices, which is the unique rooted tree in \mathbb{T}_{n+1}

of height ≤ 1 (equality holds unless $n = 0$). Note that $C_1 = S_0 = \circ$, the singleton tree.

By using the recurrence formula (3.3), we can calculate ϕ_T for chains and shrubs as follows. Consider the generating functions

$$\begin{aligned} c(x) &= \sum_{n=1}^\infty \phi_{C_n} x^n, \\ s(x) &= \sum_{n=0}^\infty \phi_{S_n} \frac{x^n}{n!}. \end{aligned}$$

Then we have:

Corollary 3.5. *The generating functions $c(x)$ and $s(x)$ are given by*

(3.5) (a) $c(x) = \ln(1 + x),$

(3.6) (b) $s(x) = \frac{x}{e^x - 1}.$

In particular, we have $\phi_{C_n} = (-1)^{n-1} \frac{1}{n}$ for all $n \geq 1$ and $\phi_{S_n} = b_n$, where b_0, b_1, b_2, \dots are the Bernoulli numbers³ defined by $\frac{x}{e^x - 1} = \sum_{n=0}^\infty b_n \frac{x^n}{n!}$.

Proof. (a) According to (3.3) we have

$$c(x) = x - \sum_{n=2}^\infty \left(\sum_{k=2}^n \frac{1}{k!} \sum_{\substack{\vec{e}=(e_1,\dots,e_{k-1}) \in E(C_n)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \phi_{T_{\vec{e},2}} \cdots \phi_{T_{\vec{e},k}} \right) x^n.$$

Noting that $v(C_n) = n$ and each $T_{\vec{e},j}$ is also a path:

$$\begin{aligned} &= x - \sum_{n=2}^\infty \sum_{k=2}^n \frac{1}{k!} \sum_{\substack{(i_1,\dots,i_k) \in \mathbb{N}^k \\ i_1 + \dots + i_k = n}} \prod_{j=1}^k \phi_{C_{i_j}} x^{i_j} \\ &= x - \sum_{n=2}^\infty \left(\sum_{k=2}^n \frac{1}{k!} (\text{coefficient of } x^n \text{ in } c(x)^k) \right) x^n \\ &= x - \sum_{n=2}^\infty \left(\text{coefficient of } x^n \text{ in } \sum_{k=2}^n \frac{1}{k!} c(x)^k \right) x^n \\ &= x - \sum_{n=2}^\infty \left(\text{coefficient of } x^n \text{ in } \sum_{k=2}^\infty \frac{1}{k!} c(x)^k \right) x^n \\ &= x - \sum_{n=2}^\infty \frac{1}{n!} c(x)^n \\ &= x - (e^{c(x)} - c(x) - 1). \end{aligned}$$

³This indexing and signage differs from an alternate definition of the Bernoulli numbers as the sequence B_1, B_2, \dots defined by

$$\frac{x}{e^x - 1} = 1 - \frac{1}{2}x + \sum_{n=1}^\infty (-1)^{n-1} \frac{B_n}{(2n)!} x^{2n}.$$

Thus the relationship is $B_n = (-1)^{n-1} b_{2n}$ for $n \geq 1$.

Solving for $c(x)$ in the equation $c(x) = x - (e^{c(x)} - c(x) - 1)$ gives (3.5).

(b) Again by (3.3) we have

$$s(x) = 1 - \sum_{n=1}^{\infty} \left(\sum_{k=2}^{v(S_n)} \frac{1}{k} \sum_{\substack{\vec{e}=(e_1,\dots,e_{k-1}) \in E(S_n)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \phi_{T_{\vec{e},2}} \cdots \phi_{T_{\vec{e},k}} \right) \frac{x^n}{n!}.$$

Noting that $v(S_n) = n + 1$ and precisely one $T_{\vec{e},j}$ is a shrub with all others being singletons:

$$\begin{aligned} &= 1 - \sum_{n=1}^{\infty} \left(\sum_{k=2}^{n+1} \frac{1}{k!} (k-1)! \binom{n}{k-1} \phi_{S_{n-k+1}} \right) \frac{x^n}{n!} \\ (3.7) \quad &= 1 - x^{-1} \sum_{n=1}^{\infty} \sum_{k=2}^{n+1} \phi_{S_{n-k+1}} \frac{x^{n-k+1}}{(n-k+1)!} \frac{x^k}{k!} \\ &= 1 - x^{-1} \left(\sum_{r=0}^{\infty} \phi_{S_r} \frac{x^r}{r!} \right) \left(\sum_{s=2}^{\infty} \frac{x^s}{s!} \right) \\ (3.8) \quad &= 1 - x^{-1} s(x) (e^x - x - 1). \end{aligned}$$

Solving for $s(x)$ in the equation $s(x) = 1 - x^{-1} s(x) (e^x - x - 1)$ gives (3.6). \square

3.3. Polynomial Coefficients $\psi_T(t)$ of Formal Flow. Let us first recall the formal flow $F_t = \exp(tA) \cdot z$ and some of its properties. See [Z] for more details.

Definition 3.6. Given an indeterminate t , define the system $F_t \in \mathbb{C}[t][[z]]^n$ by

$$(3.9) \quad F_t = \exp(tA) \cdot z.$$

It is called the formal flow generated by F .

It is easy to verify that $F_t \in \mathbb{C}[t][[z]]^n$. Therefore, a specialization $t = \alpha$, for any $\alpha \in \mathbb{C}$ (or α in any \mathbb{C} -algebra), makes sense. According to Proposition (3.1), setting $t = 1$ in F_t recovers F .

The following proposition shows that t behaves as an exponent for F .

Proposition 3.7. *Let t and s be indeterminates. Then*

$$F_{s+t} = F_t \circ F_s.$$

Hence setting $t = n$ in F_t , for $n \in \mathbb{N}$, gives the n -fold composition $F \circ \cdots \circ F$, and setting $t = -n$ gives the n -fold composition $F^{-1} \circ \cdots \circ F^{-1}$ of the formal inverse. In particular,

$$F_t|_{t=-1} = F^{-1}.$$

Proof. We have

$$\begin{aligned} F_{s+t} &= \exp((s+t)A) \cdot z = \exp(sA + tA) \cdot z \\ &= \exp(sA) \cdot \exp(tA) \cdot z = \exp(sA) \cdot F_t. \end{aligned}$$

We use the fact that $\exp(sA)$ is a \mathbb{C} -algebra automorphism of $\mathbb{C}[s, t][[z]]$ (see Remark 3.2):

$$\begin{aligned} &= F_t(\exp(sA \cdot z)) = F_t(F_s) \\ &= F_t \circ F_s. \end{aligned}$$

\square

Thus F_t can be viewed as the “formal t -th power of F ”.

The system F_t can be expressed in terms of the tree expressions \mathcal{P}_T as follows:

Theorem 3.8. *There exists a unique set of polynomials $\{\psi_T(t) \in \mathbb{Q}[t]\}$ indexed by the set of rooted trees $T \in \mathbb{T}$ such that*

$$(3.10) \quad F_t = z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T.$$

These polynomials are given by the formula

$$(3.11) \quad \psi_T(t) = \sum_{k=1}^{v(T)} \frac{t^k}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \phi_{T_{\vec{e},2}} \cdots \phi_{T_{\vec{e},k}}.$$

(Again we must interpret the $k = 1$ summand as ϕ_T .)

Proof. According to Theorem 3.4 the D-Log of F is given by $a = \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T$. Hence we have $A = aD = \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T D = \sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T$ (see Definition 2.7). Therefore,

$$\begin{aligned} F_t &= \exp(tA) \cdot z = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \cdot z \\ &= z + \sum_{k=1}^{\infty} \frac{t^k}{k!} \left(\sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T \right)^k \cdot z \\ &= z + \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_k} \mathcal{D}_{T_k}] \cdot z \\ &= z + \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_{k-1}} \mathcal{D}_{T_{k-1}}] [\phi_{T_k} \mathcal{P}_{T_k}]. \end{aligned}$$

Now we apply Corollary 2.10 to the $k \geq 2$ summands:

$$\begin{aligned} &= z + \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_{T \in \mathbb{T}} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \cdots \phi_{T_{\vec{e},k}} \mathcal{P}_T \\ &= z + \sum_{T \in \mathbb{T}} \left(\sum_{k=1}^{v(T)} \frac{t^k}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_{k-1}) \in E(T)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{\vec{e},1}} \cdots \phi_{T_{\vec{e},k}} \right) \mathcal{P}_T. \end{aligned}$$

This gives the desired result. The uniqueness of ψ_T follows from applying stable linear independence (Corollary 2.6) to each power of t in (3.10). \square

Lemma 3.9. *For any $T \in \mathbb{T}$, we have*

- (1) if T is the singleton, we have $\psi_T(t) = t$.
- (2) $\psi_T(0) = 0$.
- (3) $\psi_T(1) = \begin{cases} 1 & \text{if } v(T) = 1, \\ 0 & \text{if } v(T) \geq 2. \end{cases}$
- (4) $\psi'_T(0) = \phi_T$.

Proof. All statements above follow immediately from (3.11), except the assertion $\psi_T(1) = 0$ when $v(T) \geq 2$, which is exactly (3.4). \square

Forests. The formula (3.11) defines a unique polynomial $\psi_T(t)$ for each rooted tree T . A *forest* is the disjoint union of finitely many rooted trees. We extend the definitions of ϕ_P and $\psi_P(t)$ to any forest P as follows:

Definition 3.10. For any forest P that is the disjoint union of rooted trees T_1, \dots, T_k , we define ϕ_P to be ϕ_{T_1} if $k = 1$ and 0 otherwise. Define $\psi_P(t) = \prod_{i=1}^k \psi_{T_i}(t)$.

Lemma 3.11. Let T be a rooted tree with $v(T) \geq 2$. For any proper rooted subtree T' of T we have

$$(3.12) \quad \psi_{T \setminus T'}(t) = \sum_{k=1}^{v(T)-1} \frac{t^k}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_k) \in E(T)^k \\ e_1 \succ \dots \succ e_k \\ T_{\vec{e}, k+1} = T'}} \phi_{T_{\vec{e}, 1}} \phi_{T_{\vec{e}, 2}} \cdots \phi_{T_{\vec{e}, k}}.$$

Proof. Let $T^{[j]}$ ($j = 1, 2, \dots, d$) be the connected components of $T \setminus T'$, and let e_j^0 be the edge of T that connects $T^{[j]}$ with T' . Note that from fixed sequences $e_{j,1} \succ e_{j,2} \succ \cdots \succ e_{j,k_j} \in E(T^{[j]})$ with $k_1 + k_2 + \cdots + k_d = k - d$, appended by the edges e_j^0 , we can get $\binom{k}{(k_1+1), (k_2+1), \dots, (k_d+1)} = \frac{k!}{(k_1+1)!(k_2+1)! \cdots (k_d+1)!}$ different sequences $e_1 \succ e_2 \succ \cdots \succ e_k \in E(T)$ such that $T_{k+1} = T'$. Therefore,

$$\begin{aligned} & \sum_{k=1}^{v(T)-1} \frac{t^k}{k!} \sum_{\substack{\vec{e}=(e_1, \dots, e_k) \in E(T)^k \\ e_1 \succ \dots \succ e_k \\ T_{\vec{e}, k+1} = T'}} \phi_{T_{\vec{e}, 1}} \phi_{T_{\vec{e}, 2}} \cdots \phi_{T_{\vec{e}, k}} \\ &= \sum_{k=1}^{v(T)-1} \frac{t^k}{k!} \sum_{\substack{(k_1, \dots, k_d) \in \mathbb{N}^d \\ k_1 + k_2 + \cdots + k_d = k - d}} \frac{k!}{(k_1 + 1)!(k_2 + 1)! \cdots (k_d + 1)!} \\ & \quad \cdot \prod_{j=1}^d \sum_{\substack{\vec{e}_j=(e_{j,1}, \dots, e_{j,k_j}) \in E(T^{[j]})^{k_j} \\ e_{j,1} \succ \dots \succ e_{j,k_j}}} \phi_{T_{e_{j,1}}} \phi_{T_{e_{j,2}}} \cdots \phi_{T_{e_{j,k_j+1}}} \\ &= \prod_{j=1}^d \sum_{k_j=0}^{v(T^{[j]})-1} \frac{t^{k_j+1}}{(k_j + 1)!} \sum_{\substack{\vec{e}_j=(e_{j,1}, \dots, e_{j,k_j}) \in E(T^{[j]})^{k_j} \\ e_{j,1} \succ \dots \succ e_{j,k_j}}} \phi_{T_{e_{j,1}}} \phi_{T_{e_{j,2}}} \cdots \phi_{T_{e_{j,k_j+1}}} \\ &= \psi_{T^{[1]}}(t) \psi_{T^{[2]}}(t) \cdots \psi_{T^{[d]}}(t). \end{aligned}$$

The last equality follows from (3.11). \square

The lemma above allows us to prove the following theorem. If we let \emptyset be the empty tree and define $\mathcal{P}_{\emptyset} = z$, then Theorem 3.8 can be seen as the special case $S = \emptyset$ of the theorem below.

Theorem 3.12. *For any rooted tree S , we have*

(3.13)
$$\exp(tA) \cdot \mathcal{P}_S = \mathcal{P}_S + \sum_{T \in \mathbb{T}} \left(\sum_{\substack{T' \leq T \\ T' \cong S}} \psi_{T \setminus T'}(t) \right) \mathcal{P}_T.$$

Proof.

$$\begin{aligned} \exp(tA) \cdot \mathcal{P}_S &= \sum_{k=0}^\infty \frac{t^k}{k!} A^k \cdot \mathcal{P}_S \\ &= \sum_{k=0}^\infty \frac{t^k}{k!} \left(\sum_{T \in \mathbb{T}} \phi_T \mathcal{D} \right)^k \cdot \mathcal{P}_S \\ &= \mathcal{P}_S + \sum_{k=1}^\infty \frac{t^k}{k!} \sum_{(T_1, \dots, T_k) \in \mathbb{T}^k} [\phi_{T_1} \mathcal{D}_{T_1}] \cdots [\phi_{T_k} \mathcal{D}_{T_k}] \cdot \mathcal{P}_S. \end{aligned}$$

Apply Corollary 2.9, substituting $Y_{T_i}^{(i)} = \phi_{T_i}$:

$$\begin{aligned} &= \mathcal{P}_S + \sum_{k=1}^\infty \frac{t^k}{k!} \sum_{\substack{T \in \mathbb{T} \\ v(T) \geq 1}} \sum_{\substack{\vec{e} = (e_1, \dots, e_{k-1}) \in E(T)^k \\ e_1 \succ \cdots \succ e_k \\ T_{\vec{e}, k+1} \cong S}} \phi_{T_{\vec{e}, 1}} \cdots \phi_{T_{\vec{e}, k}} \mathcal{P}_T \\ &= \mathcal{P}_S + \sum_{\substack{T \in \mathbb{T} \\ v(T) \geq 1}} \left(\sum_{k=1}^{v(T)-1} \frac{t^k}{k!} \sum_{\substack{\vec{e} = (e_1, \dots, e_k) \in E(T)^k \\ e_1 \succ \cdots \succ e_k \\ T_{\vec{e}, k+1} \cong S}} \phi_{T_{\vec{e}, 1}} \cdots \phi_{T_{\vec{e}, k}} \right) \mathcal{P}_T \\ &= \sum_{T \in \mathbb{T}} \left(\sum_{\substack{T' \leq T \\ T' \cong S}} \psi_{T \setminus T'}(t) \right) \mathcal{P}_T. \end{aligned}$$

The last equality follows from Lemma 3.11. □

Proposition 3.13. *For any rooted tree T , we have*

(a)

(3.14)
$$\psi'_T(t) = \phi_T + \sum_{e \in E(T)} \phi_{T_{e,1}} \psi_{T_{e,2}}(t),$$

(b)

(3.15)
$$\psi'_T(t) = \phi_T + \sum_{S < T} \phi_S \psi_{T \setminus S}(t)$$

or, in other words,

(3.16)
$$\begin{aligned} \psi'_T(t) &= \psi'_T(0) + \sum_{e \in E(T)} \psi'_{T_{e,1}}(0) \psi_{T_{e,2}}(t) \\ &= \psi'_T(0) + \sum_{S < T} \psi'_S(0) \psi_{T \setminus S}(t). \end{aligned}$$

Proof. (a) Applying the chain rule and Theorem 3.8, we have

$$\begin{aligned}
 \frac{\partial}{\partial t} F_t &= \frac{\partial}{\partial t} (\exp(tA) \cdot z) \\
 &= A \cdot \exp(tA) \cdot z \\
 &= \left(\sum_{T \in \mathbb{T}} \phi_T \mathcal{D}_T \right) \cdot \left(z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T \right) \\
 &= \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T + \sum_{(T_1, T_2) \in \mathbb{T}^2} \phi_{T_1} \psi_{T_2}(t) \mathcal{D}_{T_1} \mathcal{P}_{T_2}.
 \end{aligned}$$

Applying Corollary 2.10 with $k = 2$, setting $Y_T^{(1)} = \phi_T$, $Y_T^{(2)} = \psi_T(t)$ for all $T \in \mathbb{T}$:

$$\begin{aligned}
 &= \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T + \sum_{T \in \mathbb{T}} \sum_{e \in E(T)} \phi_{T_{e,1}} \psi_{T_{e,2}}(t) \mathcal{P}_T \\
 &= \sum_{T \in \mathbb{T}} \left(\phi_T + \sum_{e \in E(T)} \phi_{T_{e,1}} \psi_{T_{e,2}}(t) \right) \mathcal{P}_T.
 \end{aligned}$$

But we also have, by Theorem 3.8,

$$(3.17) \quad \frac{\partial}{\partial t} F_t = \sum_{T \in \mathbb{T}} \psi'_T(t) \mathcal{P}_T.$$

Comparing the coefficient of \mathcal{P}_T , and appealing to stable linear independence—specifically, Corollary 2.6—we get (3.14). (We use the fact that polynomial functions that agree at all $\alpha \in \mathbb{C}$ must be equal.)

(b)

$$\begin{aligned}
 \frac{\partial}{\partial t} F_t &= \frac{\partial}{\partial t} \exp(tA) \cdot z \\
 &= A \cdot \exp(tA) \cdot z \\
 &= \exp(tA) \cdot A \cdot z \\
 &= \exp(tA) \cdot a.
 \end{aligned}$$

Applying Theorem 3.4:

$$\begin{aligned}
 &= \exp(tA) \cdot \sum_{S \in \mathbb{T}} \phi_S \mathcal{P}_S \\
 &= \sum_{S \in \mathbb{T}} \phi_S \exp(tA) \cdot \mathcal{P}_S.
 \end{aligned}$$

Applying Theorem 3.12:

$$\begin{aligned} &= \sum_{S \in \mathbb{T}} \phi_S \mathcal{P}_S + \sum_{S \in \mathbb{T}} \phi_S \sum_{T \in \mathbb{T}} \left(\sum_{\substack{T' < T \\ T' \cong S}} \psi_{T \setminus T'}(t) \right) \mathcal{P}_T \\ &= \sum_{T \in \mathbb{T}} \phi_T \mathcal{P}_T + \sum_{T \in \mathbb{T}} \left(\sum_{S < T} \phi_S \psi_{T \setminus T'}(t) \right) \mathcal{P}_T \\ &= \sum_{T \in \mathbb{T}} \left(\phi_T + \sum_{S < T} \phi_S \psi_{T \setminus T'}(t) \right) \mathcal{P}_T. \end{aligned}$$

Comparing this with (3.17), and again employing Corollary 2.6, we get (3.15). \square

An interesting consequence of the proposition above is the following recurrence formula for ϕ_T in terms of the number of the leaves of T .

Proposition 3.14. *For any rooted tree T , suppose that the root rt_T has d children, i.e., $d = |rt_T^+|$. Then*

$$(3.18) \quad \sum_{r=0}^{l(T)} \sum_{\substack{\{v_1, v_2, \dots, v_r\} \subseteq L(T) \\ v_1, v_2, \dots, v_r \text{ distinct}}} \phi_{T \setminus \{v_1, v_2, \dots, v_r\}} = \delta_{d,1} \phi_{T \setminus \{rt_T\}}.$$

Proof. From (3.14), setting $t = 1$, we get

$$(3.19) \quad \psi'_T(1) = \begin{cases} \phi_T + \phi_{T \setminus \{rt_T\}} & \text{if } d = 1, \\ \phi_T & \text{if } d \geq 2 \end{cases}$$

since, by Lemma 3.9, $\psi_{T_2}(1) = 0$ except when T_2 is the singleton. From (3.15), setting $t = 1$, we get

$$(3.20) \quad \psi'_T(1) = \phi_T + \sum_{r=1}^{l(T)} \sum_{\substack{\{v_1, v_2, \dots, v_r\} \subseteq L(T) \\ v_1, v_2, \dots, v_r \text{ distinct}}} \phi_{T \setminus \{v_1, v_2, \dots, v_r\}}$$

since $\psi_{T \setminus S}(1) = 0$ except when $T \setminus S$ is the disjoint union of finitely many singletons. Comparing (3.19) and (3.20) gives (3.18). \square

Before leaving this subsection, we will do some calculations on the polynomials $\psi_T(t)$ for the chains C_n and shrubs S_n .

Consider the generating functions $\mathcal{C}(t, x) = \sum_{n=0}^\infty \psi_{C_n}(t) x^n$ (set $\psi_{C_0}(t) = 1$) and $\mathcal{S}(t, x) = \sum_{n=0}^\infty \psi_{S_n}(t) \frac{x^n}{n!}$.

Corollary 3.15. *The generating functions $\mathcal{C}(t, x)$ and $\mathcal{S}(t, x)$ are given by*

(a)

$$(3.21) \quad \mathcal{C}(t, x) = \exp(t \ln(1 + x)) = (1 + x)^t$$

or, in other words,

$$(3.22) \quad \psi_{C_n}(t) = \binom{t}{n} = \frac{t(t-1) \cdots (t-n+1)}{n!}.$$

(b)

$$(3.23) \quad \mathcal{S}(t, x) = \frac{e^{xt} - 1}{e^x - 1}.$$

Proof. (a) By Theorem 3.8 and Corollary 3.5 we have

$$\begin{aligned} \mathcal{C}(t, x) &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{v(C_n)} \frac{t^k}{k!} \sum_{\substack{\bar{e}=(e_1, \dots, e_{k-1}) \in E(C_n)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{e,1}} \phi_{T_{e,2}} \cdots \phi_{T_{e,k}} x^n \\ &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^n \frac{t^k}{k!} \sum_{\substack{(m_1, m_2, \dots, m_k) \in (\mathbb{Z}^+)^k \\ m_1 + m_2 + \dots + m_k = n}} \frac{(-1)^{m_1}}{m_1} \frac{(-1)^{m_2}}{m_2} \cdots \frac{(-1)^{m_k}}{m_k} x^n \\ &= e^{t(-x + \frac{x^2}{2} - \dots + \frac{(-x)^m}{m} + \dots)} \\ &= \exp(t \ln(1+x)). \end{aligned}$$

(b) Similarly, we have

$$\mathcal{S}(t, x) = \sum_{n=0}^{\infty} \sum_{k=1}^{v(S_n)} \frac{t^k}{k!} \sum_{\substack{\bar{e}=(e_1, \dots, e_{k-1}) \in E(S_n)^{k-1} \\ e_1 \succ \dots \succ e_{k-1}}} \phi_{T_{e,1}} \phi_{T_{e,2}} \cdots \phi_{T_{e,k}} \frac{x^n}{n!}.$$

Noting that all but one of $\phi_{T_{e,2}}, \dots, \phi_{T_{e,k}}$ are singletons, the remaining one being S_{n-k+1} :

$$\begin{aligned} &= \sum_{n=0}^{\infty} \sum_{k=1}^{n+1} \frac{t^k}{k!} (k-1)! \binom{n}{k-1} b_{n-k+1} \frac{x^n}{n!} \\ &= x^{-1} \sum_{n=0}^{\infty} \sum_{k=1}^{n+1} \frac{(xt)^k}{k!} b_{n-k+1} \frac{x^{n-k+1}}{(n-k+1)!}. \end{aligned}$$

Replacing n by $n-1$:

$$\begin{aligned} &= x^{-1} \sum_{n=1}^{\infty} \sum_{k=1}^n \frac{(xt)^k}{k!} b_{n-k} \frac{x^{n-k}}{(n-k)!} \\ &= x^{-1} (e^{xt} - 1) \frac{x}{e^x - 1} \\ (3.24) \quad &= \frac{e^{xt} - 1}{e^x - 1}. \end{aligned}$$

□

Remark 3.16. The formulas of Corollary 3.15 can also be easily derived from Theorem 4.2 in the next section. But we think the calculations above are more intriguing.

4. THE MAIN THEOREM

4.1. Main Theorem on $\psi_T(t)$. In the last section, we defined the polynomial $\psi_T(t)$, for each rooted tree T (see Theorem 3.8). For each rooted forest P , i.e., the disjoint union of finitely many rooted trees T_i ($i = 1, 2, \dots, k$), we also defined ψ_P

(see Definition 3.10). Recalling from §2.1 the definition of a rooted subtree, we are now ready to prove the following main theorem.

Theorem 4.1. *Let t and s be indeterminates. For $T \in \mathbb{T}$ we have*

$$(4.1) \qquad \psi_T(t+s) = \psi_T(t) + \psi_T(s) + \sum_{T' < T} \psi_{T \setminus T'}(t) \psi_{T'}(s)$$

where the last sum runs over all proper rooted subtrees T' of T .

Proof. Clearly $\exp((t+s)A) \cdot z = \exp(tA) \cdot \exp(sA) \cdot z$. So we have

$$\begin{aligned} z + \sum_{T \in \mathbb{T}} \psi_T(t+s) \mathcal{P}_T &= \exp((t+s)A) \cdot z = \exp(tA) \cdot \exp(sA) \cdot z \\ &= \exp(tA) \cdot \left(z + \sum_{T \in \mathbb{T}} \psi_T(s) \mathcal{P}_T \right) \\ &= \exp(tA) \cdot z + \exp(tA) \cdot \left(\sum_{T \in \mathbb{T}} \psi_T(s) \mathcal{P}_T \right) \\ &= z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \psi_T(s) (\exp(tA) \cdot \mathcal{P}_T). \end{aligned}$$

Applying Theorem 3.12 to $\exp(tA) \cdot \mathcal{P}_T$:

$$\begin{aligned} &= z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \psi_T(s) \left(\mathcal{P}_T + \sum_{S \in \mathbb{T}} \left(\sum_{\substack{S' < S \\ S' \cong T}} \psi_{S \setminus S'}(t) \right) \mathcal{P}_S \right) \\ &= z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \psi_T(s) \mathcal{P}_T + \sum_{S \in \mathbb{T}} \left(\sum_{T \in \mathbb{T}} \sum_{\substack{S' < S \\ S' \cong T}} \psi_{S \setminus S'}(t) \psi_{T'}(s) \right) \mathcal{P}_S \\ &= z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \psi_T(s) \mathcal{P}_T + \sum_{S \in \mathbb{T}} \sum_{S' < S} \psi_{S \setminus S'}(t) \psi_{S'}(s) \mathcal{P}_S. \end{aligned}$$

Replacing S by T in the last summation:

$$= z + \sum_{T \in \mathbb{T}} \psi_T(t) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \psi_T(s) \mathcal{P}_T + \sum_{T \in \mathbb{T}} \sum_{T' < T} \psi_{T \setminus T'}(t) \psi_{T'}(s) \mathcal{P}_T.$$

According to Corollary 2.6 and an easy specialization argument, the theorem follows by comparing the coefficients of \mathcal{P}_T in the above equation. □

The *difference polynomial* of $g(t) \in \mathbb{C}[T]$ is defined to be the polynomial $\Delta g(t) = g(t+1) - g(t)$. The following special case of the theorem above, which gives a formula for the difference polynomial of $\psi_T(t)$, is most useful to us.

Theorem 4.2. *For any tree T with $v(T) \geq 2$, we have*

$$\begin{aligned} \Delta \psi_T(t) &= \psi_{T_1}(t) \psi_{T_2}(t) \cdots \psi_{T_d}(t) \\ &= \psi_{T \setminus \{rt_T\}}(t) \end{aligned}$$

where T_i , $i = 1, 2, \dots, d$ are the connected components of $T \setminus \{rt_T\}$.

Proof. This follows from Theorem 4.1 by setting $s = 1$ in (4.1) and appealing to Lemma 3.9, which says that $\psi_T(1) = 0$ unless T is the singleton, in which case $\psi_T(1) = 1$. \square

Theorem 4.3. *For any tree T with $v(T) \geq 2$, we have*

(a)

$$(4.2) \quad \Delta\psi_T(t) = \sum_{r=1}^{l(T)} \sum_{\substack{\{v_1, v_2, \dots, v_r\} \subseteq L(T) \\ v_1, v_2, \dots, v_r \text{ distinct}}} \psi_{T \setminus \{v_1, v_2, \dots, v_r\}}(t).$$

(b)

$$(4.3) \quad \psi_{T \setminus \{rt_T\}}(t) = \sum_{r=1}^{l(T)} \sum_{\substack{\{v_1, v_2, \dots, v_r\} \subseteq L(T) \\ v_1, v_2, \dots, v_r \text{ distinct}}} \psi_{T \setminus \{v_1, v_2, \dots, v_r\}}(t)$$

where T_i , $i = 1, 2, \dots, d$ are the connected components of $T \setminus \{rt_T\}$.

Proof. Clearly, (b) follows from (a) and Theorem 4.2. For (a), switch t and s and set $s = 1$ in 4.1 to get

$$\psi_T(t+1) = \psi_T(t) + \psi_T(1) + \sum_{T' < T} \psi_{T \setminus T'}(1) \psi_{T'}(t).$$

By Lemma 3.9, we have $\psi_T(1) = 0$ and $\psi_{T \setminus T'}(1) = 0$, unless $T \setminus T'$ is a disjoint union of singletons, in which case $\psi_{T \setminus T'}(1) = 1$. Therefore,

$$\psi_T(t+1) - \psi_T(t) = \sum_{r=1}^{l(T)} \sum_{\substack{\{v_1, v_2, \dots, v_r\} \subseteq L(T) \\ v_1, v_2, \dots, v_r \text{ distinct}}} \psi_{T \setminus \{v_1, v_2, \dots, v_r\}}(t)$$

as desired. \square

4.2. Algorithm for $\psi_T(t)$. From Theorem 4.2 we get the following algorithm for computing $\psi_T(t)$. Here, for $h(t) \in \mathbb{C}[t]$, $\Delta^{-1}h(t)$ is defined to be the unique polynomial $g(t) \in \mathbb{C}[t]$ such that $\Delta g(t) = h(t)$ and $g(0) = 0$.

Algorithm. For any fixed rooted tree T , we sign a polynomial $N_v(t)$ to each vertex v of T as follows:

- (1) For each leaf v of T , set $N_v(t) = t$.
- (2) For any other vertex v of T , define $N_v(t)$ inductively starting from the highest level by $N_v(t) = \Delta^{-1}(N_{v_1}(t)N_{v_2}(t) \cdots N_{v_k}(t))$, where v_j , $j = 1, 2, \dots, k$, are the distinct children of v .

Then for each vertex v of T , $N_v(t) = \psi_{T_v^+}(t)$, where T_v^+ is the subtree of T rooted at the vertex v . In particular, we have $\psi_T(t) = N_{rt_T}(t)$. \square

The following example applies this algorithm to the shrubs S_n to show that the polynomials $\psi_{S_n}(t)$ are closely related to the Bernoulli polynomials $B_n(t)$ defined by $\frac{xe^{tx}}{e^x - 1} = \sum_{n=0}^{\infty} B_n(t) \frac{x^n}{n!}$. (Compare this with (b) of Corollary 3.15.)

Example 4.4. Let v_1, \dots, v_n be the leaves of the shrub S_n . Following the algorithm, we first assign the polynomial t to each leaf v_i . The next step in the algorithm gives

$$(4.4) \quad \psi_{S_n}(t) = \Delta^{-1}(t^n).$$

One of the fundamental properties of the Bernoulli polynomials $B_n(t)$ is

$$(4.5) \qquad \Delta B_n(t) = B_n(t+1) - B_n(t) = nt^{n-1},$$

and from this and the fact that Δ commutes with $\frac{d}{dt}$ one easily derives

$$(4.6) \qquad \frac{d}{dt} B_{n+1}(t) = (n+1)B_n(t).$$

From (4.5) and (4.6) we get

$$(4.7) \qquad \Delta^{-1}(t^n) = \int_0^t B_n(u)du = \frac{B_{n+1}(t) - B_{n+1}(0)}{n+1}.$$

Putting together equations (4.4) and (4.7), we obtain this relationship between $\psi_{S_n}(t)$ and $B_{n+1}(t)$,

$$\psi_{S_n}(t) = \int_0^t B_n(u)du = \frac{B_{n+1}(t) - B_{n+1}(0)}{n+1}.$$

4.3. Combinatorial Property of $\psi_T(t)$. After the main part of this work was done, Professor John Shareshian pointed out to us that the polynomial $\psi_T(t)$ for rooted trees coincides with the strict order polynomial $\bar{\Omega}(P, t)$ for finite posets (partial ordered sets) P in combinatorics (see Chapters 3 and 4 in [St1]). We first recall the polynomial $\bar{\Omega}(P, t)$ associated with a finite poset, and then we show that, when P is the poset of the set $V(T)$ of vertices of a rooted tree T with the natural partial order induced by ancestry (the root being the unique smallest element), we have $\psi_T(t) = \bar{\Omega}(P, t)$.

Any rooted tree corresponds in this way to a unique finite poset, and a finite poset P corresponds to a rooted tree precisely when it satisfies these two criteria:

- (1) P has a unique smallest element, and
- (2) any interval in P is totally ordered.

For any $n \in \mathbb{N}$, the chain C_n gives the totally ordered poset with n elements. (We can view it as the set $\{1, 2, \dots, n\}$ with the natural order of the positive integers.) For any poset P , we say that a map $f : P \rightarrow C_n$ is *strict order-preserving* if $f(a) < f(b)$ in C_n whenever $a < b$ in P . It is well known that there exists a unique polynomial $\bar{\Omega}(P, t)$ such that $\bar{\Omega}(P, n)$ equals the number of strict order-preserving maps f from P to C_n for all $n \in \mathbb{N}$. This, then, is the theorem shown to us by John Shareshian.

Theorem 4.5. *For any rooted tree T , we have*

$$(4.8) \qquad \psi_T(t) = \bar{\Omega}(T, t)$$

(where, on the right, T is viewed as a finite poset as described above).

Proof. It is obvious that when T is the singleton, $\bar{\Omega}(T, t) = t$. Hence it is enough to show that $\bar{\Omega}(T, t)$ also satisfies the recursion formula of Theorem 4.2. More precisely, we will show that, in the notation of Theorem 4.2, we have

$$\Delta \bar{\Omega}(T, n) = \bar{\Omega}(T, n+1) - \bar{\Omega}(T, n) = \bar{\Omega}(T_1, n) \bar{\Omega}(T_2, n) \cdots \bar{\Omega}(T_d, n)$$

for any $n \in \mathbb{N}$.

Note that $\Delta \bar{\Omega}(T, n)$ equals the number of strict order-preserving maps f from T to $C_{n+1} = \{1, 2, \dots, n+1\}$ such that $f(\text{rt}_T) = 1$. But this number is also the same as the number of strict order-preserving maps g from $T \setminus \{\text{rt}_T\}$ to C_n , which is $\bar{\Omega}(T_1, n) \bar{\Omega}(T_2, n) \cdots \bar{\Omega}(T_d, n)$. □

Remark 4.6. It is interesting that the strict order polynomial $\bar{\Omega}(T, t)$ for the finite posets induced by rooted trees T can be defined in a totally different way, namely, according to the formula (3.11) of Theorem 3.8. In fact, this realization of the strict order polynomial can be generalized to an arbitrary finite poset P . This generalization and its consequences will be discussed in the upcoming paper [SWZ].

5. SOME APPLICATIONS

For a formal automorphism $F = (F_1, F_2, \dots, F_n) = z + H$ of the form identity plus higher, we give a restatement and new proof of the tree formula for the formal inverse first proved in [BCW] and [W2].

Theorem 5.1. *For any rooted tree T , we have $\psi_T(-1) = (-1)^{v(T)}$. Hence the formal inverse F^{-1} of F is given by⁴*

$$(5.1) \quad F^{-1} = z + \sum_{T \in \mathbb{T}} (-1)^{v(T)} \mathcal{P}_T.$$

Proof. The formula (5.1) follows from $\psi_T(-1) = (-1)^{v(T)}$ by Proposition 3.7 and Theorem 3.8.

It is well known in combinatorics (see [St1]) that the strict order polynomials satisfy $\bar{\Omega}(T, -1) = (-1)^{v(T)}$, from which the result follows, in light of Theorem 4.5. For completeness, we give a direct proof here.

We use the mathematical induction on $v(T)$. The case for $v(T) = 1$ is trivial. ($\psi_T(t) = t$ in this case.) Suppose $v(T) \geq 2$. By Theorem 4.2, setting $t = -1$, we have

$$\psi_T(0) - \psi_T(-1) = \psi_{T_1}(-1)\psi_{T_2}(-1) \cdots \psi_{T_d}(-1)$$

where T_i , $i = 1, 2, \dots, d$ are the connected components of $T \setminus \{\text{rt}_T\}$. We have $\psi_T(0) = 0$, and by induction we may assume the theorem holds for T_1, \dots, T_d . Hence

$$\psi_T(-1) = -(-1)^{v_{T_1} + v_{T_2} + \cdots + v_{T_d}} = (-1)^{v(T)}.$$

□

It is known that the Jacobian conjecture (see [BCW] for a statement of this famous problem) is equivalent to the assertion that

$$(5.2) \quad \sum_{T \in \mathbb{T}_N} \mathcal{P}_T = 0$$

for $N \gg 0$ whenever H is a homogeneous polynomial system (of degree ≥ 2) and the Jacobian determinant $|(D_j F_i)|$ is (everywhere) nonzero. In fact, this follows from Theorem 5.1, since when H is homogeneous the polynomials $\sum_{T \in \mathbb{T}_N} \mathcal{P}_T$, for fixed N , are the homogeneous summands of F^{-1} (see Remark 2.2). When H is homogeneous, the condition $|(D_j F_i)| = 1$ is known to be equivalent to the nilpotence of the Jacobian matrix $JH = (D_j H_i)$ (see [BCW]). Thus the following result presents an intriguing statement for comparison.

⁴The formula as given in [BCW] and [W2] did not include the factor $(-1)^{v(T)}$. It appears here because of our choice in writing $F = z + H$ instead of $F = z - H$.

Proposition 5.2. *Assume H is homogeneous of degree ≥ 2 . For any rooted tree, let $h_{T,k}$ be the number of vertices of height k . Suppose that $(JH)^k = 0$. Then*

$$(5.3) \qquad \sum_{T \in \mathbb{T}_N} h_{T,m} \mathcal{P}_T = 0$$

for any $N \in \mathbb{N}$ and $m \geq k$.

Proof. Suppose that $\deg H = d \geq 2$. It follows from Euler’s formula that $JH \cdot (z^t) = (dH)^t$, from which we get $\frac{1}{d}(JH)^k \cdot (z^t) = (JH)^{k-1} \cdot H^t = 0$. (Here the superscript t denotes transpose, converting a row to a column so that the matrix multiplications make sense.) For any integer $m \geq 1$, a straightforward calculation shows that the chain C_m has the property $\mathcal{P}_{C_m} = JH^{m-1} \cdot H^t$. Therefore, $\mathcal{P}_{C_m} = 0$ for $m \geq k$, and we have

$$0 = \exp(-A) \cdot \mathcal{P}_{C_m}.$$

By Theorem 3.12, setting $S = C_m$ and $t = -1$ in (3.13):

$$= \sum_{T \in \mathbb{T}} \left(\sum_{\substack{T' \leq T \\ T' \cong C_m}} \psi_{T \setminus T'}(-1) \right) \mathcal{P}_T.$$

By Theorem 5.1:

$$\begin{aligned} &= \sum_{T \in \mathbb{T}} \left(\sum_{\substack{T' \leq T \\ T' \cong C_m}} (-1)^{v(T \setminus T')} \right) \mathcal{P}_T \\ &= \sum_{T \in \mathbb{T}} (-1)^{v(T)-m} h_{T,m} \mathcal{P}_T. \end{aligned}$$

In particular, for any $N \in \mathbb{N}$, we have

$$\sum_{T \in \mathbb{T}_N} (-1)^{N-m} h_{T,m} \mathcal{P}_T = (-1)^{N-m} \sum_{T \in \mathbb{T}_N} h_{T,m} \mathcal{P}_T = 0,$$

which gives (5.3). □

The proposition above shows that, for a fixed homogeneous polynomial system H , the polynomials \mathcal{P}_T are in some sense quite linearly dependent on each other.

Finally, let us point out that the formal flow F_t gives a formal flow between F and the identity map id , i.e., $F_t|_{t=1} = F$ and $F_t|_{t=0} = \text{id}$, having the additional properties $F_t(0) = 0$ and $JF_t(0) = I_n$. It is an open question in complex analysis whether, for any local analytic map F , such an analytic flow exists. The usual approach to this question is to show that F is linearizable, i.e., it is conjugate to a linear map. But when F is linearizable the question is still open, even for the one-variable case. (There are many partial results for this problem.) So it is of interest that the formal solution to this question is given by the very clean formula (3.10) of Theorem 3.8. But the question of when F_t is locally convergent is still open.

REFERENCES

- [A] S. S. Abhyankar, *Lectures in algebraic geometry*, Notes by Chris Christensen, Purdue Univ., 1974.
- [BCW] H. Bass, E. Connell, and D. Wright, *The Jacobian conjecture, reduction of degree and formal expansion of the inverse*. Bull. Amer. Math. Soc. **7**, (1982), 287–330. MR **83k**:14028. Zbl.539.13012.
- [CMTWW] C. C.-A. Cheng, J. H. McKay, J. Towber, S. S.-S. Wang, and D. Wright, *Reversion of power series and the extended Raney coefficients*, Trans. Amer. Math. Soc. **349** (1997), 1769–1782. MR **97h**:13018 Zbl.868.13019.
- [Ge] I. M. Gessel, *A combinatorial proof of the multivariable Lagrange inversion formula*, J. Combin. Theory Ser. A, **45** (1987), 178–195. MR **88h**:05011 Zbl.651.05009.
- [Go] I. J. Good, *Generalizations to several variables of Lagrange's expansion, with applications to stochastic processes*, Proc. Cambridge Philos. Soc. **56** (1960), 367–380. MR **23**:A352 Zbl.135.18802.
- [HS] M. Haiman and W. Schmitt, *Incidence algebra antipodes and Lagrange inversion in one and several variables*, J. Combin. Theory Ser. A, **50** (1989), 172–185. MR **90f**:05005 Zbl.747.05007.
- [Ja] C. G. J. Jacobi, *De resolutione aequationum per series infinitas*, J. Reine Angew. Math. **184** (1830), 257–286.
- [Jo] S. A. Joni, *Lagrange inversion in higher dimensions and umbral operators*, Linear and Multilinear Algebra **6** (1978) 111–122. MR **58**:10485 Zbl.395.05005.
- [R] G. N. Raney, *Functional composition patterns and power series reversion*, Trans. Amer. Math. Soc. **94** (1960), 441–451. MR **22**:5584 Zbl.131.14.
- [St1] Richard P. Stanley, *Enumerative Combinatorics I*, Cambridge University Press, 1997. MR **98a**:05001 Zbl.945.05006.
- [St2] Richard P. Stanley, *Enumerative Combinatorics II*, Cambridge University Press, 1999. MR **2000k**:05026 Zbl.928.05001.
- [Sh] J. Shareshian, *Personal communication*.
- [SWZ] J. Shareshian, D. Wright and W. Zhao, *A New Realization of Order Polynomials*. In Preparation.
- [W1] D. Wright, *Formal inverse expansion and the Jacobian Conjecture*, J. Pure Appl. Algebra, **48** (1987), 199–219. MR **89b**:13008 Zbl.666.12017.
- [W2] D. Wright, *The tree formulas for reversion of power series*, J. Pure Appl. Algebra, **57** (1989) 191–211. MR **90d**:13008 Zbl.672.13010.
- [Z] W. Zhao, *Exponential formulas for the Jacobians and Jacobian matrices of analytic maps*, J. Pure Appl. Algebra, **166** (2002) 321–336. MR **2002i**:14059

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN ST. LOUIS, ST. LOUIS, MISSOURI 63130-4899

E-mail address: wright@einstein.wustl.edu

DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN ST. LOUIS, ST. LOUIS, MISSOURI 63130-4899

E-mail address: zhao@math.wustl.edu

A GENERALIZATION OF TIGHT CLOSURE AND MULTIPLIER IDEALS

NOBUO HARA AND KEN-ICHI YOSHIDA

ABSTRACT. We introduce a new variant of tight closure associated to any fixed ideal \mathfrak{a} , which we call \mathfrak{a} -tight closure, and study various properties thereof. In our theory, the annihilator ideal $\tau(\mathfrak{a})$ of all \mathfrak{a} -tight closure relations, which is a generalization of the test ideal in the usual tight closure theory, plays a particularly important role. We prove the correspondence of the ideal $\tau(\mathfrak{a})$ and the multiplier ideal associated to \mathfrak{a} (or, the adjoint of \mathfrak{a} in Lipman's sense) in normal \mathbb{Q} -Gorenstein rings reduced from characteristic zero to characteristic $p \gg 0$. Also, in fixed prime characteristic, we establish some properties of $\tau(\mathfrak{a})$ similar to those of multiplier ideals (e.g., a Briançon-Skoda-type theorem, subadditivity, etc.) with considerably simple proofs, and study the relationship between the ideal $\tau(\mathfrak{a})$ and the F -rationality of Rees algebras.

INTRODUCTION

The notion of tight closure, introduced by Hochster and Huneke [HH1] more than a decade ago, has emerged as a powerful new tool in commutative algebra. Tight closure gives remarkably simple characteristic p proofs of several results that were not thought to be particularly related, e.g., that rings of invariants of linearly reductive groups acting on regular rings are Cohen–Macaulay, that the integral closure of the n th power of an n -generator ideal of a regular ring is contained in the ideal (the Briançon–Skoda theorem), and so on. Also, the notions of F -regular and F -rational rings are defined via tight closure, and they turned out to correspond to log terminal and rational singularities, respectively ([Ha1], [HW], [MS], [Sm1]). This result is generalized to the correspondence of test ideals and multiplier ideals ([Ha2], [Sm2]), both of which play very important roles in the tight closure theory and birational algebraic geometry, respectively.

The test ideal of a ring R of characteristic p , denoted by $\tau(R)$, is the annihilator ideal of all tight closure relations of R . On the other hand, the notion of multiplier ideals has several variants. Originally, a multiplier ideal was defined analytically for a given plurisubharmonic function on a complex analytic manifold [N]. This is reformulated in the algebro-geometric setting (in characteristic zero) in terms of resolution of singularities and discrepancy divisors ([Ei], [La]). Actually, two types of multiplier ideals are defined in this setting, that is, the multiplier ideal $\mathcal{J}(D)$ associated to a \mathbb{Q} -divisor D and the multiplier ideal $\mathcal{J}(\mathfrak{a})$ associated to an ideal \mathfrak{a} .

Received by the editors August 20, 2002 and, in revised form, December 19, 2002.

2000 *Mathematics Subject Classification.* Primary 13A35, 14B05.

Both authors are partially supported by a Grant-in-Aid for Scientific Research, Japan.

The latter is also defined by Lipman [Li] in a more algebraic context and is called the “adjoint ideal”.

Precisely speaking, the multiplier ideal that is proved to correspond to the test ideal $\tau(R)$ is the one associated to the trivial divisor $D = 0$ or the unit ideal $\mathfrak{a} = R$, which defines the non-log-terminal locus of $\text{Spec } R$. In most applications, however, the usefulness of multiplier ideals is obtained by considering multiplier ideals associated to various divisors or ideals; see, e.g., [Ei], [La], [Li]. Thus we are tempted to define a sort of tight closure operation and a “test ideal” associated to any given \mathbb{Q} -divisor or ideal.

In this paper, we introduce a generalization of tight closure, which we call \mathfrak{a} -tight closure, associated to an ideal \mathfrak{a} , study various properties (including the relationship with multiplier ideals), and give some applications of \mathfrak{a} -tight closure. (In [T], the reader can find an attempt to generalize tight closure in the other direction, that is, Δ -tight closure associated to a \mathbb{Q} -divisor Δ ; see also [HW].) Actually, given an ideal \mathfrak{a} of a Noetherian ring R of characteristic $p > 0$, we define the \mathfrak{a} -tight closure $I^{*\mathfrak{a}}$ of an ideal $I \subseteq R$ to be the ideal consisting of all elements $z \in R$ for which there exists an element $c \in R$ not in any minimal prime ideal such that

$$cz^{p^e} \mathfrak{a}^{p^e} \subseteq I^{[p^e]}$$

for all $e \gg 0$, where $I^{[p^e]}$ is the ideal generated by the p^e th powers of elements of I ; see Definition 1.1 and also Definition 6.1 for further generalization to “rational coefficients”. We then define the ideal $\tau(\mathfrak{a})$ of R to be the unique largest ideal such that $\tau(\mathfrak{a})I^{*\mathfrak{a}} \subseteq I$ for all ideals $I \subseteq R$. So, in the case where $\mathfrak{a} = R$ is the unit ideal, the \mathfrak{a} -tight closure $I^{*\mathfrak{a}} = I^{*R}$ is equal to the tight closure I^* in the usual sense, and the ideal $\tau(\mathfrak{a}) = \tau(R)$ is nothing but the test ideal [HH1].

There are many similarities between the usual tight closure and \mathfrak{a} -tight closure. For example, the existence of \mathfrak{a} -test elements (Definition 1.6) is proved quite similarly as in the case for the usual test elements. But there does exist a difference as well: We require a “closure” operation to satisfy the property that, once the operation is performed, the obtained closure does not change if one performs the operation twice or more, and tight closure satisfies this property, namely, $(I^*)^* = I^*$. However, it happens that $(I^{*\mathfrak{a}})^{*\mathfrak{a}}$ is strictly larger than $I^{*\mathfrak{a}}$, and so, \mathfrak{a} -tight closure is not in fact a closure operation. Similarly, unlike the usual test ideal $\tau(R)$, the ideal $\tau(\mathfrak{a})$ is no longer equal to the one generated by \mathfrak{a} -test elements if $\mathfrak{a} \subsetneq R$.

In spite of the apparent disadvantage mentioned above, we find many more advantages in the circle of ideas involving \mathfrak{a} -tight closure. The significance of the ideal $\tau(\mathfrak{a})$ is witnessed by the following theorem, which ensures the expected correspondence of $\tau(\mathfrak{a})$ and the multiplier ideal $\mathcal{J}(\mathfrak{a})$; see also Theorem 6.7.

Theorem 3.4. *Let R be a normal \mathbb{Q} -Gorenstein local ring essentially of finite type over a field, and let \mathfrak{a} be a nonzero ideal. Assume that $\mathfrak{a} \subseteq R$ is reduced from characteristic zero to characteristic $p \gg 0$, together with a log resolution of singularities $f: X \rightarrow Y = \text{Spec } R$ such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ is invertible. Then*

$$\tau(\mathfrak{a}) = H^0(X, \mathcal{O}_X([K_X - f^*K_Y] - Z)).$$

Note that, by definition, the multiplier ideal $\mathcal{J}(\mathfrak{a})$ in characteristic zero takes just the same form as the right-hand side of the above equality. So one can think of the right-hand side as a reduction modulo p of the multiplier ideal in characteristic zero. This theorem generalizes the main results of [Ha2] and [Sm2], and is proved essentially in a similar way as in those papers, with a little more effort.

The usefulness of multiplier ideals is tied up with vanishing theorems in characteristic zero such as the Nadel vanishing theorem [N], which fail in characteristic $p > 0$. For example, Lipman used these tools to establish an improved version of the Briançon–Skoda theorem, which asserts that, for any ideal \mathfrak{a} of a regular local ring generated by r elements, one has $\mathcal{J}(\mathfrak{a}^{n+r-1}) \subseteq \mathfrak{a}^n$ for all $n \geq 0$ [Li]; see also [La]. We will prove in Theorem 2.1 that the corresponding statement $\tau(\mathfrak{a}^{n+r-1}) \subseteq \mathfrak{a}^n$ holds true in characteristic $p > 0$. Taking into account the correspondence in Theorem 3.4, we see that this gives another proof of Lipman’s result. An advantage of our prime characteristic proof here is that it is quite elementary and simple (like the original tight closure proof of Briançon–Skoda [HH1]) and does not depend on desingularization or vanishing theorems.

We shall take a look at the organization of this paper, which we hope gives further confirmation of the usefulness of \mathfrak{a} -tight closure.

After discussing basic properties of \mathfrak{a} -tight closure and the ideal $\tau(\mathfrak{a})$ in Section 1, we give three fundamental applications of \mathfrak{a} -tight closure in Section 2. The first one is the modified Briançon–Skoda theorem via \mathfrak{a} -tight closure mentioned above. Second, we study the relationship between \mathfrak{a} -tight closure and tight integral closure defined by Hochster [Ho2], and rephrase the F-rationality criterion of Rees algebras obtained in [HWY1] in terms of \mathfrak{a} -tight closure; see Theorem 2.7 and Corollary 2.9. This enables us to give an interesting characterization of regular local rings as the third application. Namely, we prove in Theorem 2.15 that the regularity of a d -dimensional local ring (R, \mathfrak{m}) is characterized by the property that $\tau(\mathfrak{m}^{d-1}) = R$. This is considered an analog of the fact that the weak F-regularity of R is characterized by the property that $\tau(R) = R$.

Section 3 is devoted to proving the equality $\tau(\mathfrak{a}) = \mathcal{J}(\mathfrak{a})$ in Theorem 3.4, which holds true in the situation reduced from characteristic zero to characteristic $p \gg 0$. We note that the containment $\tau(\mathfrak{a}) \subseteq \mathcal{J}(\mathfrak{a})$ essentially holds true in any fixed characteristic $p > 0$; cf. Proposition 3.8.

In Section 4, we establish properties of ideals $\tau(\mathfrak{a})$ in fixed characteristic $p > 0$ similar to those of multiplier ideals $\mathcal{J}(\mathfrak{a})$ in characteristic zero, namely, the restriction theorem (Theorem 4.1), the subadditivity (Theorem 4.5), and a description of ideals $\tau(\mathfrak{a})$ in the toric case (Theorem 4.8); see [DEL], [How], [La] for the results proved for multiplier ideals. Again in light of Theorem 3.4, we can also say that the results in this section give new prime characteristic proofs of the geometric statements for multiplier ideals in characteristic zero, although we work with the Frobenius map in fixed characteristic $p > 0$.

In Section 5, we explore the behavior of the ideal $\tau(I)$ for an \mathfrak{m} -primary ideal I of a Gorenstein local domain (R, \mathfrak{m}) of characteristic $p > 0$, from the viewpoint of the Rees algebra $\mathbf{R}(I) = R[It]$ via Theorem 2.7. As a main result of this section, we prove in Theorem 5.1 that if $\mathbf{R}(I)$ is F-rational, then its graded canonical module is described as $\omega_{\mathbf{R}(I)} = \bigoplus_{n \geq 1} \tau(I^n)t^n$. In particular, the equality $\tau(I) = \mathcal{J}(I)$ holds if $\mathbf{R}(I)$ is F-rational, and the converse is also true in the case of a two-dimensional rational double point. (It should be noted that the ideals $\tau(I)$ and $\mathcal{J}(I)$ may disagree in fixed positive characteristic.) Comparing Theorem 5.1 with Hyry’s results ([Hy1], [Hy2]), we can also deduce various results for $\tau(I)$.

In Section 6, we extend the notions of \mathfrak{a} -tight closure and the ideal $\tau(\mathfrak{a})$ to those with “rational coefficients”, and generalize some results discussed in the previous sections to the case of rational coefficients. Although we have no explicit applica-

tions of this generalization at the moment, we include this section for future reference, because recent applications of multiplier ideals involve rational coefficients successfully ([ELS], [La]).

Before going ahead, we review part of the notation and basic notions of the tight closure theory. We keep it minimal to avoid overlap with Section 1. The reader is referred to Hochster and Huneke [HH1]–[HH3] and Huneke [Hu] for the full development of the theory.

Notation and basic notions. Throughout this paper all rings are Noetherian commutative rings with unity. For a ring R , we denote by R° the set of elements of R that are not in any minimal prime ideal. We will often work over a field of characteristic $p > 0$. In this case we always use the letter q for a power p^e of p .

Let R be a Noetherian ring of characteristic $p > 0$. For an ideal I of R and a power q of p , we denote by $I^{[q]}$ the ideal generated by the q th powers of elements of I . The *tight closure* I^* of I is defined to be the ideal consisting of all elements $z \in R$ for which there exists an element $c \in R^\circ$ such that $cz^q \in I^{[q]}$ for all large $q = p^e$. (Tight closure is also defined for a submodule of a module; cf. Section 1.)

We say that R is *weakly F -regular* if every ideal I of R is tightly closed, that is, $I^* = I$. A local ring (R, \mathfrak{m}) is said to be *F -rational* if every ideal generated by a system of parameters of R is tightly closed. In general, we say that R is *F -rational* (resp. *F -regular*) if all of its local rings are F -rational (resp. weakly F -regular).

Let $F: R \rightarrow R$ be the Frobenius map, that is, the ring homomorphism sending $z \in R$ to $z^p \in R$. The ring R viewed as an R -module via the e -times iterated Frobenius map $F^e: R \rightarrow R$ is denoted by eR . We say that R is *F -finite* if 1R is a finitely generated R -module. If R is reduced, then $F^e: R \rightarrow {}^eR$ is identified with the natural inclusion map $R \hookrightarrow R^{1/q}$. An F -finite reduced ring R is said to be *strongly F -regular* if for every element $c \in R^\circ$, there exists a power $q = p^e$ such that the inclusion map $c^{1/q}R \hookrightarrow R^{1/q}$ splits as an R -module homomorphism.

1. DEFINITION AND BASIC PROPERTIES OF \mathfrak{a} -TIGHT CLOSURE

Let R be a Noetherian ring of characteristic $p > 0$ and let M be an R -module. For each $e \in \mathbb{N}$, we define $\mathbb{F}^e(M) = \mathbb{F}_R^e(M) := {}^eR \otimes_R M$ and regard it as an R -module by the action of $R = {}^eR$ from the left. Then we have the induced e -times iterated Frobenius map $F^e: M \rightarrow \mathbb{F}^e(M)$. The image of $z \in M$ via this map is denoted by $z^q := F^e(z) \in \mathbb{F}^e(M)$. For an R -submodule N of M , we denote by $N_M^{[q]}$ the image of the induced map $\mathbb{F}^e(N) \rightarrow \mathbb{F}^e(M)$.

Definition 1.1. Let \mathfrak{a} be an ideal of a Noetherian ring R of characteristic $p > 0$ and let $N \subseteq M$ be R -modules. The *\mathfrak{a} -tight closure* of N in M , denoted by $N_M^{*\mathfrak{a}}$, is defined to be the submodule of M consisting of all elements $z \in M$ for which there exists $c \in R^\circ$ such that

$$cz^q \mathfrak{a}^q \subseteq N_M^{[q]}$$

for all large $q = p^e$. The \mathfrak{a} -tight closure of an ideal $I \subseteq R$ is just defined by $I^{*\mathfrak{a}} = I_R^{*\mathfrak{a}}$.

Remark 1.2. (1) In the case where $\mathfrak{a} = R$ is the unit ideal, the \mathfrak{a} -tight closure $N_M^{*\mathfrak{a}} = N_M^{*R}$ is nothing but the (usual) tight closure N_M^* as defined in [HH1]. However, unlike the usual tight closure, it may happen that $(N_M^{*\mathfrak{a}})^{*\mathfrak{a}}$ is strictly larger than $N_M^{*\mathfrak{a}}$; see Remark 1.4 (1). In this sense \mathfrak{a} -tight closure is not an “honest” closure operation in general.

(2) It seems significant to extend the definition to “rational coefficients”, if we take into account the relationship with multiplier ideals; see [La], [T] and Sections 3 and 4. Namely, given nonnegative $t \in \mathbb{Q}$ and $\mathfrak{a} \subseteq R$, $N \subseteq M$ as in Definition 1.1, we can define the $t \cdot \mathfrak{a}$ -tight closure $N_M^{*t, \mathfrak{a}}$ of N in M and generalize some results for \mathfrak{a} -tight closure to those for $t \cdot \mathfrak{a}$ -tight closure. We treat this issue in Section 6.

We collect some basic properties of \mathfrak{a} -tight closure in the following. The proofs are easy and are left to the reader.

Proposition 1.3. *Let \mathfrak{a} and \mathfrak{b} denote ideals of a Noetherian ring R of characteristic $p > 0$ and let L and N denote submodules of an R -module M .*

- (1) $N \subseteq N_M^{*\mathfrak{a}}$ and $N_M^{*\mathfrak{a}}/N \cong 0_{M/N}^{*\mathfrak{a}}$.
- (2) If $L \subseteq N$, then $L_M^{*\mathfrak{a}} \subseteq N_M^{*\mathfrak{a}}$.
- (3) $N_M^{*\mathfrak{a}\mathfrak{b}} \subseteq (N_M^{*\mathfrak{a}} : \mathfrak{b})_M$. Moreover, if \mathfrak{b} is a principal ideal, then the equality $N_M^{*\mathfrak{a}\mathfrak{b}} = (N_M^{*\mathfrak{a}} : \mathfrak{b})_M$ holds.
- (4) If $\mathfrak{b} \subseteq \mathfrak{a}$, then $N_M^{*\mathfrak{a}} \subseteq N_M^{*\mathfrak{b}}$. Moreover, if $\mathfrak{a} \cap R^\circ \neq \emptyset$ and if \mathfrak{b} is a reduction of \mathfrak{a} , then the equality $N_M^{*\mathfrak{a}} = N_M^{*\mathfrak{b}}$ holds.

Remark 1.4. (1) It follows that $N_M^* \subseteq N_M^{*\mathfrak{a}} \subseteq (N_M^* : \mathfrak{a})_M$ from (3) and (4) of Proposition 1.3. If \mathfrak{a} is a principal ideal, then the equality on the right occurs, and $(N_M^{*\mathfrak{a}})_M^{*\mathfrak{a}} = (N_M^* : \mathfrak{a}^2)_M$ is strictly larger than $N_M^{*\mathfrak{a}} = (N_M^* : \mathfrak{a})_M$ in general.

(2) The colon-capturing property [HH1, Section 7] says that parameters behave like a regular sequence modulo tight closure. Namely, if $x_1, \dots, x_{i+1} \in R$ are parameters, then under a mild assumption, $(x_1, \dots, x_i) :_R x_{i+1} \subseteq (x_1, \dots, x_i)^*$. Since \mathfrak{a} -tight closure contains the usual tight closure, this remains true if we replace the usual tight closure by \mathfrak{a} -tight closure. In Proposition 1.5 below we slightly improve this colon-capturing property for \mathfrak{a} -tight closure using the existence of a test element. See [HH1], [HH2] for the definition and detailed study of test elements, and see also Definition 1.6 for a generalization to the notion of \mathfrak{a} -test elements.

Proposition 1.5. *Let R be an equidimensional reduced excellent ring of characteristic $p > 0$ and let \mathfrak{a} be an ideal. Then for any parameters x_1, \dots, x_n in R ,*

$$(x_1, \dots, x_{n-1})^{*\mathfrak{a}} : x_n \subseteq (x_1, \dots, x_{n-1})^{*\mathfrak{a}}.$$

Proof. Actually, we can prove even more. Namely, let $I, J \subseteq R$ be ideals generated by monomials in parameters x_1, \dots, x_n , and let $K \subseteq R$ be the “expected” answer for $I : J$, that is, the monomial ideal that would be equal to $I : J$ if x_1, \dots, x_n formed a regular sequence. (Note that $I = K = (x_1, \dots, x_{n-1})$ and $J = (x_n)$ in our case.) Then we will show that

$$I^{*\mathfrak{a}} : J \subseteq K^{*\mathfrak{a}}.$$

Let $z \in I^{*\mathfrak{a}} : J$. Then there is a $c \in R^\circ$ such that $cz^q \mathfrak{a}^q \subseteq I^{[q]} : J^{[q]}$ for $q = p^e \gg 0$. Since the “expected” answer for $I^{[q]} : J^{[q]}$ is $K^{[q]}$, we have that $I^{[q]} : J^{[q]} \subseteq (K^{[q]})^*$ by colon-capturing of the usual tight closure [HH1]. So, for a test element $d \in R^\circ$, one has that $(cd)z^q \mathfrak{a}^q \subseteq K^{[q]}$ for $q = p^e \gg 0$, which means that $z \in K^{*\mathfrak{a}}$, as required. \square

Definition 1.6. Let R be a Noetherian ring of characteristic $p > 0$ and let \mathfrak{a} be an ideal of R . We say that an element $c \in R^\circ$ is an \mathfrak{a} -test element if $cz^q \mathfrak{a}^q \subseteq I^{[q]}$ for all $q = p^e$ whenever $z \in I^{*\mathfrak{a}}$.

In the case where $\mathfrak{a} = R$ is the unit ideal, an \mathfrak{a} -test element is nothing but a test element in the usual sense [HH1]. In [HH2] it is proved that a test element exists in nearly every ring of interest, for example, in excellent reduced local rings [HH2, Theorem 6.1]. We can show that an \mathfrak{a} -test element also does.

Theorem 1.7. *Let R be a reduced Noetherian ring of characteristic $p > 0$, let $c \in R^\circ$, and assume that one of the following conditions holds:*

- (1) *R is F -finite and the localized ring R_c is strongly F -regular;*
- (2) *R is an excellent local ring and R_c is Gorenstein and F -regular.*

Then some power c^n of c is an \mathfrak{a} -test element for all ideals $\mathfrak{a} \subseteq R$.

Here we prove the above theorem under assumption (1) only, according to the method of [HH0]. The case of assumption (2) is reduced to the F -finite case by the machinery of “ T -construction” used in [HH2]. We do not include the argument involving this reduction process, because it is somewhat long but essentially the same as that for the usual tight closure [HH2, Section 6].

To prove the theorem in the F -finite case we need the following lemma from [HH0, Remark 3.2], in which it is implicit that the exponent n of c may be independent of the choice of d .

Lemma 1.8. *Let R be an F -finite reduced Noetherian ring of characteristic $p > 0$. If the localization R_c of R at an element $c \in R^\circ$ is strongly F -regular, then there exists an integer $n \geq 0$, depending only on R and c , satisfying the following property: For any $d \in R^\circ$, there exist a power q' of p and an R -linear map $\phi: R^{1/q'} \rightarrow R$ sending $d^{1/q'}$ to c^n .*

Proof of Theorem 1.7 in case (1). We will show that c^n in Lemma 1.8 is an \mathfrak{a} -test element for every $\mathfrak{a} \subseteq R$. Given any ideal I , any $z \in I^{*\mathfrak{a}}$ and any power q of p , it is enough to show that $c^n z^q \mathfrak{a}^q \subseteq I^{[q]}$. Since $z \in I^{*\mathfrak{a}}$, there exists $d \in R^\circ$ such that $dz^Q \mathfrak{a}^Q \subseteq I^{[Q]}$ for every Q . Then by Lemma 1.8, there exist q' and $\phi: R^{1/q'} \rightarrow R$ sending $d^{1/q'}$ to c^n . Since $dz^{qq'} (\mathfrak{a}^q)^{[q']} \subseteq dz^{qq'} \mathfrak{a}^{qq'} \subseteq I^{[qq']}$, one has

$$d^{1/q'} z^q \mathfrak{a}^q R^{1/q'} \subseteq I^{[q]} R^{1/q'},$$

and applying ϕ gives $c^n z^q \mathfrak{a}^q \subseteq I^{[q]}$, as required. □

Proposition-Definition 1.9. *Let R be a Noetherian ring of characteristic $p > 0$ and let \mathfrak{a} be an ideal of R . Let $E = \bigoplus_{\mathfrak{m}} E_R(R/\mathfrak{m})$, the direct sum, taken over all maximal ideals \mathfrak{m} of R , of the injective envelopes of the residue fields R/\mathfrak{m} . Then the following ideals are equal to each other, and we denote them by $\tau(\mathfrak{a})$:*

- i) $\bigcap_M \text{Ann}_R(0_M^{*\mathfrak{a}})$, where M runs through all finitely generated R -modules;
- ii) $\bigcap_{M \subseteq E} \text{Ann}_R(0_M^{*\mathfrak{a}})$, where M runs through all finitely generated R -submodules of E .

If R is locally approximately Gorenstein (e.g., if R is excellent and reduced [Ho1]), then the following ideal is also equal to $\tau(\mathfrak{a})$:

- iii) $\bigcap_{I \subseteq R} (I : I^{*\mathfrak{a}})$, where I runs through all ideals of R .

Proof. The proof is the same as that for the usual tight closure. See [HH1, Proposition 8.23] for the equality of i) and ii), and [HH1, Proposition 8.25] for the equality of ii) and iii). □

Remark 1.10. In the case where $\mathfrak{a} = R$ is the unit ideal, $\tau(\mathfrak{a}) = \tau(R)$ is called the test ideal. In this case, $\tau(R) \cap R^\circ$ is equal to the set of test elements of R , and this justifies the name “test ideal”. But the name “ \mathfrak{a} -test ideal” for $\tau(\mathfrak{a})$ is somewhat misleading if $\mathfrak{a} \neq R$, because $\tau(\mathfrak{a}) \cap R^\circ$ is not equal to the set of \mathfrak{a} -test elements in general.

The following basic properties of the ideal $\tau(\mathfrak{a})$ follow from Proposition 1.3. See Theorem 2.1 for a generalization of the latter half of (1).

Proposition 1.11. *Let R be a Noetherian ring of characteristic $p > 0$ and let \mathfrak{a} and \mathfrak{b} denote ideals of R .*

- (1) $\tau(\mathfrak{a})\mathfrak{b} \subseteq \tau(\mathfrak{a}\mathfrak{b})$. Moreover, if \mathfrak{b} is a principal ideal of a complete local ring, then $\tau(\mathfrak{a})\mathfrak{b} = \tau(\mathfrak{a}\mathfrak{b})$.
- (2) If $\mathfrak{b} \subseteq \mathfrak{a}$, then $\tau(\mathfrak{b}) \subseteq \tau(\mathfrak{a})$. Moreover, if $\mathfrak{a} \cap R^\circ \neq \emptyset$ and if \mathfrak{b} is a reduction of \mathfrak{a} , then $\tau(\mathfrak{b}) = \tau(\mathfrak{a})$.
- (3) If R admits a test element and if $\mathfrak{a} \cap R^\circ \neq \emptyset$, then $\tau(\mathfrak{a}) \cap R^\circ \neq \emptyset$.
- (4) If R is weakly F -regular, then $\mathfrak{a} \subseteq \tau(\mathfrak{a})$. Moreover, if \mathfrak{a} is an ideal of pure height one, then $\mathfrak{a} = \tau(\mathfrak{a})$.

Proof. The first half of (1) is immediate from Proposition 1.3 (3). To prove the second half, let (R, \mathfrak{m}) be a complete local ring and let \mathfrak{b} be principal. Then by the Matlis duality, $\text{Ann}_E(\tau(\mathfrak{a}))$ is equal to the union of $0_M^{\mathfrak{a}}$ taken over all finitely generated submodules M of $E = E_R(R/\mathfrak{m})$. So, if $z \in \text{Ann}_E(\tau(\mathfrak{a})\mathfrak{b})$, then there exists a finitely generated submodule $M \subset E$ such that $z \in (0_M^{\mathfrak{a}} : \mathfrak{b})_E$. Replacing M by $M + Rz \subset E$, one has $z \in (0_M^{\mathfrak{a}} : \mathfrak{b})_M = 0_M^{\mathfrak{a}\mathfrak{b}}$ by Proposition 1.3 (3). Hence

$$\tau(\mathfrak{a}\mathfrak{b}) = \bigcap_{M \subseteq E} \text{Ann}_R(0_M^{\mathfrak{a}} : \mathfrak{b})_M = \text{Ann}_R(\text{Ann}_E(\tau(\mathfrak{a})\mathfrak{b})) = \tau(\mathfrak{a})\mathfrak{b}.$$

(2) follows from Proposition 1.3 (4), and (3) and the first half of (4) from $\tau(R)\mathfrak{a} \subseteq \tau(\mathfrak{a})$. As for the second half of (4), it suffices to show the following claim, since weakly F -regular rings are normal.

Claim 1.11.1. *If R is normal and \mathfrak{a} is an ideal of pure height one, then $\tau(\mathfrak{a}) \subseteq \mathfrak{a}$.*

To prove the claim, considering a primary decomposition of \mathfrak{a} , we may assume, without loss of generality, that \mathfrak{a} is a primary ideal such that $\mathfrak{p} = \sqrt{\mathfrak{a}}$ is a height one prime ideal. Then, since $R_{\mathfrak{p}}$ is a discrete valuation ring, we can choose $b \in \mathfrak{a}$ such that $\mathfrak{a}R_{\mathfrak{p}} = bR_{\mathfrak{p}}$. Then $bR : \mathfrak{a} \subseteq \mathfrak{a}^{*a}$. Indeed, if $z \in bR : \mathfrak{a}$, then $z^q \mathfrak{a}^q \subseteq b^q R \subseteq \mathfrak{a}^{[q]}$ for all $q = p^e$, so that $z \in \mathfrak{a}^{*a}$. It now follows from $bR : \mathfrak{a} \not\subseteq \mathfrak{p}$ that $\mathfrak{a}^{*a} \not\subseteq \mathfrak{p}$. Since \mathfrak{a} is \mathfrak{p} -primary, we have $\tau(\mathfrak{a}) \subseteq \mathfrak{a} : \mathfrak{a}^{*a} = \mathfrak{a}$, as claimed. \square

Proposition 1.12 (cf. [B], [HH1, Proposition 4.12]). *Let $R \subseteq S$ be a pure ring extension of Noetherian rings of characteristic $p > 0$ such that $R^\circ \subseteq S^\circ$. Then for any ideal \mathfrak{a} of R , one has $\tau(\mathfrak{a}S) \cap R \subseteq \tau(\mathfrak{a})$.*

Proof. For a finitely generated R -module M , the natural map $M = M \otimes_R R \rightarrow M \otimes_R S$ is injective by the purity of $R \subseteq S$. Since $R^\circ \subseteq S^\circ$, we see easily that $0_M^{\mathfrak{a}} \subseteq 0_{M \otimes_R S}^{\mathfrak{a}S}$ via the inclusion map $M \hookrightarrow M \otimes_R S$. Hence, if $c \in \tau(\mathfrak{a}S) \cap R$, then c kills $0_M^{\mathfrak{a}}$ for all finitely generated R -modules M , so that $c \in \tau(\mathfrak{a})$. \square

By definition, the ideal $\tau(\mathfrak{a})$ is the annihilator of \mathfrak{a} -tight closure relations for all ideals or finitely generated modules. It will be very useful if $\tau(\mathfrak{a})$ is determined by

\mathfrak{a} -tight closure relations for a single ideal or a single module. Let us take a look at some cases where this is true.

Theorem 1.13. *Let (R, \mathfrak{m}) be a d -dimensional excellent normal local ring of characteristic $p > 0$, \mathfrak{a} an ideal of R , and let $J \subseteq R$ be a divisorial ideal such that the divisor class $\text{cl}(J) \in \text{Cl}(R)$ has a finite order. Then*

$$0_{H_{\mathfrak{m}}^d(J)}^{*\mathfrak{a}} = \bigcup_{M \subset H_{\mathfrak{m}}^d(J)} 0_M^{*\mathfrak{a}},$$

where M runs through all finitely generated R -submodules of $H_{\mathfrak{m}}^d(J)$. In particular, if R is \mathbb{Q} -Gorenstein, then

$$\tau(\mathfrak{a}) = \text{Ann}_R(0_E^{*\mathfrak{a}}),$$

where $E = E_R(R/\mathfrak{m}) \cong H_{\mathfrak{m}}^d(\omega_R)$.

Proof. Again the proof is the same as that for the usual tight closure,¹ but we sketch a proof according to [Sm2, Lemma 3.4], which is based on the idea of [AM, 3.1].

Let r be the order of $\text{cl}(J) \in \text{Cl}(R)$ and let $J^{(r)} = x_1 R$. We may assume, without loss of generality, that $x_1 \in \mathfrak{m}$. Then there exist $x_2 \in R$ and $0 \neq a \in J$ such that $x_2 J \subseteq aR$, and x_1, x_2 extends to a system of parameters x_1, x_2, \dots, x_d for R .

The point of the proof is that $\mathbb{F}^e(H_{\mathfrak{m}}^d(J)) \cong H_{\mathfrak{m}}^d(J^{(p^e)})$ is computed by

$$H_{\mathfrak{m}}^d(J^{(q)}) = \varinjlim R/(x_1^s J^{(q)}, x_2^s, \dots, x_d^s),$$

where the direct limit map $R/(x_1^s J^{(q)}, x_2^s, \dots, x_d^s) \rightarrow R/(x_1^{s+1} J^{(q)}, x_2^{s+1}, \dots, x_d^{s+1})$ is the multiplication by $x_1 x_2 \cdots x_d$. Then an element $\xi \in H_{\mathfrak{m}}^d(J)$ is represented by $z \bmod (x_1^s J, x_2^s, \dots, x_d^s) \in R/(x_1^s J, x_2^s, \dots, x_d^s)$ for some $z \in R$ and $s \in \mathbb{N}$, and $\xi = [z \bmod (x_1^s J, x_2^s, \dots, x_d^s)]$ is mapped to $\xi^{p^e} = [z^{p^e} \bmod (x_1^{p^e s} J^{(p^e)}, x_2^{p^e s}, \dots, x_d^{p^e s})]$ by the e -times iterated Frobenius map $F^e: H_{\mathfrak{m}}^d(J) \rightarrow H_{\mathfrak{m}}^d(J^{(p^e)})$.

Now say that $\xi \in 0_{H_{\mathfrak{m}}^d(J)}^{*\mathfrak{a}}$. Then there exists $c \in R^\circ$ such that $c\xi^q \alpha = [cz^q \alpha \bmod (x_1^{q^s} J^{(q)}, x_2^{q^s}, \dots, x_d^{q^s})] = 0$ for all $q = p^e \gg 0$ and $\alpha \in \mathfrak{a}^q$. Since \mathfrak{a}^q is a finitely generated ideal for each $q = p^e$, there exists $t_e \in \mathbb{N}$ such that $cz^q (x_1 \cdots x_d)^{t_e} \mathfrak{a}^q \subseteq (x_1^{qs+t_e} J^{(q)}, x_2^{qs+t_e}, \dots, x_d^{qs+t_e}) \subseteq (x_1^{qs+t_e+\lfloor q/r \rfloor}, x_2^{qs+t_e}, \dots, x_d^{qs+t_e})$. Then one has $cz^q \mathfrak{a}^q \subseteq (x_1^{qs+\lfloor q/r \rfloor}, x_2^{qs}, \dots, x_d^{qs})^*$ by colon-capturing. Replacing c by cc' with c' a test element and multiplying by x_1 , we see that $cx_1 z^q \mathfrak{a}^q \subseteq (x_1^{qs} J^{(q)}, x_2^{qs}, \dots, x_d^{qs})$. This gives

$$cx_1 (x_1 \cdots x_d)^q z^q \mathfrak{a}^q \subseteq (x_1^{q(s+1)} \mathfrak{a}^q, x_2^{q(s+1)}, \dots, x_d^{q(s+1)}) \subseteq (x_1^{s+1} J, x_2^{s+1}, \dots, x_d^{s+1})^{[q]}$$

for all $q = p^e \gg 0$, whence $(x_1 \cdots x_d)z \in (x_1^{s+1} J, x_2^{s+1}, \dots, x_d^{s+1})^{*\mathfrak{a}}$. Hence ξ is in the \mathfrak{a} -tight closure of zero in the cyclic (hence finitely generated) submodule of $H_{\mathfrak{m}}^d(J)$ generated by the image of $R/(x_1^{s+1} J, x_2^{s+1}, \dots, x_d^{s+1})$. \square

Discussion 1.14. In Sections 2 and 5, we will consider when the equality $\tau(\mathfrak{a}) = R$ holds. When R is a Gorenstein local ring, one can check this condition looking only at the \mathfrak{a} -tight closure of a single parameter ideal, as we will see below; cf. [FW].

Let (R, \mathfrak{m}) be a d -dimensional Cohen–Macaulay local ring, \mathfrak{a} any ideal of R , and let J be the ideal generated by a system of parameters x_1, \dots, x_d . Then

¹The proof in [Ha2, Appendix] has a minor gap at the bottom of p. 1904, although the result [Ha2, Theorem 1.8] itself and the arguments in the cited references [Mc], [Wi] are valid.

$H_m^d(R) \cong \varinjlim R/(x_1^t, \dots, x_d^t)$, and R/J and $H_m^d(R)$ have the same socle in common via the natural inclusion map $R/J \hookrightarrow H_m^d(R)$. Then $0_{H_m^d(R)}^{*\alpha} = 0$ if and only if $0_{R/J}^{*\alpha} = 0$ or, equivalently, if $J^{*\alpha} = J$. In particular, the condition that J is α -tightly closed does not depend on the choice of a parameter ideal J .

Now assume further that (R, \mathfrak{m}) is Gorenstein. Then $E = E_R(R/\mathfrak{m}) \cong H_m^d(R)$, and one sees easily that $\tau(\alpha) = \text{Ann}_R(0_{H_m^d(R)}^{*\alpha}) = \bigcap_{t \in \mathbb{N}} (x_1^t, \dots, x_d^t) : (x_1^t, \dots, x_d^t)^{*\alpha}$. Therefore $\tau(\alpha) = R$ if and only if $J^{*\alpha} = J$ for some (or equivalently, every) ideal J generated by a system of parameters. \square

As we have seen so far, the α -tight closure of the zero submodule in the injective envelope $E_R(R/\mathfrak{m})$ or the top local cohomology $H_m^d(R)$ of a local ring (R, \mathfrak{m}) plays a particularly important role. We close this section by the following proposition, which generalizes Smith's characterization of the usual tight closure of the zero submodule in $H_m^d(R)$.

Proposition 1.15 (cf. [Sm1]). *Let (R, \mathfrak{m}) be a d -dimensional excellent normal local ring of characteristic $p > 0$ and let $\alpha \subseteq R$ be an ideal such that $\alpha \cap R^\circ \neq \emptyset$. Then $0_{H_m^d(R)}^{*\alpha}$ is the unique maximal proper submodule N with respect to the property*

$$\alpha^q F^e(N) \subseteq N \text{ for all } q = p^e,$$

where $F^e: H_m^d(R) \rightarrow H_m^d(R)$ is the e -times iterated Frobenius induced on $H_m^d(R)$.

Proof. Let $c \in R^\circ$ be an element such that R_c is regular. Then \hat{R}_c is also regular by the excellence of R . Hence some power c^n of c is an α -test element and an $\alpha\hat{R}$ -test element, by Theorem 1.7. It is easy to see that c^n also works as a test element for both the α -tight closure and the $\alpha\hat{R}$ -tight closure of the zero submodule in $H_m^d(R) = H_{m\hat{R}}^d(\hat{R})$ (see the proof of Theorem 1.13). Then it follows that $0_{H_m^d(R)}^{*\alpha} = 0_{H_{m\hat{R}}^d(\hat{R})}^{*\alpha\hat{R}}$; so we may assume, without loss of generality, that R is a complete local ring.

It is easy to see that $\alpha^q F^e(0_{H_m^d(R)}^{*\alpha}) \subseteq 0_{H_m^d(R)}^{*\alpha}$ for all $q = p^e$. Also, $0_{H_m^d(R)}^{*\alpha}$ is a proper submodule of $H_m^d(R)$, because it is annihilated by $\tau(\alpha)$ by Theorem 1.13 and $\tau(\alpha) \cap R^\circ \neq \emptyset$ by Proposition 1.11 (3). To prove the maximality of $0_{H_m^d(R)}^{*\alpha}$, suppose that $N \subset H_m^d(R)$ is a proper submodule such that $\alpha^q F^e(N) \subseteq N$ for all $q = p^e$. Then the Matlis dual of the exact sequence $0 \rightarrow N \rightarrow H_m^d(R) \rightarrow H_m^d(R)/N \rightarrow 0$ is

$$0 \rightarrow [H_m^d(R)/N]^\vee \rightarrow \omega_R \rightarrow N^\vee \rightarrow 0,$$

where $[H_m^d(R)/N]^\vee$ is a nonzero submodule of ω_R . So both $[H_m^d(R)/N]^\vee$ and ω_R are torsion-free R -modules of rank 1. Therefore N^\vee is a finitely generated torsion module, so that there exists $c \in R^\circ$ such that $cN^\vee = 0$. This implies that $cN = cN^{\vee\vee} = 0$, so that $c\alpha^q F^e(N) = 0$ for all $q = p^e$. Hence $N \subseteq 0_{H_m^d(R)}^{*\alpha}$, as required. \square

2. α -TIGHT CLOSURE AND ITS APPLICATIONS

In this section we give some fundamental applications of α -tight closure.

Modified Briançon-Skoda theorem via α -tight closure. One of the important applications of tight closure theory [HH1] is a prime characteristic proof of the Briançon-Skoda theorem [BS], which was originally proved by an analytic method. Later, Lipman [Li] improved this in terms of adjoint ideals. The following is a prime characteristic analog of Lipman's "modified Briançon-Skoda" [Li, Theorem 1.4.1]; see also Remark 3.2 (1).

Theorem 2.1. *Let R be a Noetherian ring of characteristic $p > 0$. If $\mathfrak{a} \subseteq R$ is an ideal generated by r elements, then*

$$\tau(\mathfrak{a}^{n+r-1}) \subseteq \mathfrak{a}^n$$

for all $n \geq 0$. If we assume further that R is a complete local ring, then

$$\tau(\mathfrak{a}^{n+r-1}\mathfrak{b}) \subseteq \tau(\mathfrak{b})\mathfrak{a}^n$$

for all $n \geq 0$ and all ideals $\mathfrak{b} \subseteq R$.

Proof. Let $\mathfrak{b} \subseteq R$ be any ideal, M any finitely generated R -module, and suppose that $z \in (0_M^{\mathfrak{b}} : \mathfrak{a}^n)_M$. Then there exists $c \in R^\circ$ such that $cz^q(\mathfrak{a}^n)^{[q]}\mathfrak{b}^q = 0$ in $\mathbb{F}^e(M)$ for all $q = p^e \gg 0$. Since $\mathfrak{a}^{(n+r-1)q} \subseteq (\mathfrak{a}^n)^{[q]}$ by the assumption, this implies that $cz^q(\mathfrak{a}^{n+r-1}\mathfrak{b})^q = 0$ in $\mathbb{F}^e(M)$ for all $q = p^e \gg 0$, so that $z \in 0_M^{\mathfrak{a}^{n+r-1}\mathfrak{b}}$. Thus $(0_M^{\mathfrak{b}} : \mathfrak{a}^n)_M \subseteq 0_M^{\mathfrak{a}^{n+r-1}\mathfrak{b}}$ and, in particular, $\text{Ann}_M(\mathfrak{a}^n) \subseteq 0_M^{\mathfrak{a}^{n+r-1}}$. Taking the intersection of the annihilator ideals over all finitely generated R -submodules $M \subseteq E = \bigoplus_{\mathfrak{m}} E_R(R/\mathfrak{m})$, we obtain

$$\tau(\mathfrak{a}^{n+r-1}) \subseteq \bigcap_{M \subseteq E} \text{Ann}_R(\text{Ann}_M(\mathfrak{a}^n)) = \text{Ann}_R(\text{Ann}_E(\mathfrak{a}^n)) = \mathfrak{a}^n.$$

Now assume that (R, \mathfrak{m}) is a complete local ring. Then $\text{Ann}_E(\tau(\mathfrak{b})) = \bigcup_{M \subseteq E} 0_M^{\mathfrak{b}}$ by the Matlis duality, and it follows as in the latter half of Proposition 1.11 (1) that $\text{Ann}_E(\tau(\mathfrak{b})\mathfrak{a}^n) \subseteq \bigcup_{M \subseteq E} (0_M^{\mathfrak{b}} : \mathfrak{a}^n)_M \subseteq \bigcup_{M \subseteq E} 0_M^{\mathfrak{a}^{n+r-1}\mathfrak{b}}$, where the unions are taken over all finitely generated submodules M of E . Thus we conclude that

$$\tau(\mathfrak{a}^{n+r-1}\mathfrak{b}) \subseteq \bigcap_{M \subseteq E} \text{Ann}_R(0_M^{\mathfrak{b}} : \mathfrak{a}^n)_M = \text{Ann}_R(\text{Ann}_E(\tau(\mathfrak{b})\mathfrak{a}^n)) = \tau(\mathfrak{b})\mathfrak{a}^n.$$

□

Remark 2.2. The tight closure version of the Briançon–Skoda theorem [HH1, Theorem 5.4] says that if \mathfrak{a} is generated by r elements, then $\overline{\mathfrak{a}^{n+r-1}} \subseteq (\mathfrak{a}^n)^*$ for all $n \geq 0$, where $\overline{\mathfrak{b}}$ denotes the integral closure of an ideal \mathfrak{b} . This implies that $\tau(R)\overline{\mathfrak{a}^{n+r-1}} \subseteq \mathfrak{a}^n$, and in the case where the test ideal $\tau(R)$ is a strong test ideal (this is the case if R is a reduced complete local ring [Vr]), $\tau(R)\overline{\mathfrak{a}^{n+r-1}} \subseteq \tau(R)\mathfrak{a}^n$. Theorem 2.1 may be considered a slight improvement of these assertions, because $\tau(R)\overline{\mathfrak{a}^{n+r-1}} \subseteq \tau(\overline{\mathfrak{a}^{n+r-1}}) = \tau(\mathfrak{a}^{n+r-1})$ by basic properties of \mathfrak{a} -tight closure; see also Discussion 5.2.

Recently, using arguments similar to the above, the first author and S. Takagi proved a sharpened version of Theorem 2.1 [HT]; cf. [Li], [La]: if (R, \mathfrak{m}) is a complete local ring of characteristic $p > 0$ and if \mathfrak{a} is an ideal with a reduction generated by r elements, then $\tau(\mathfrak{a}^{n+r-1}) = \tau(\mathfrak{a}^{r-1})\mathfrak{a}^n$ for all $n \geq 0$.

Corollary 2.3. *Let R be a reduced excellent ring of characteristic $p > 0$ and let \mathfrak{a} be an ideal such that $\mathfrak{a} \cap R^\circ \neq \emptyset$. Then for any R -modules $N \subset M$ and any $z \in M$, the following conditions are equivalent.*

- (1) $z \in N_M^{\mathfrak{a}}$, i.e., there exists $c \in R^\circ$ such that $cz^q\mathfrak{a}^q \subseteq N_M^{[q]}$ for all $q = p^e$.
- (2) There exists $c \in R^\circ$ such that $cz^q\tau(\mathfrak{a}^q) \subseteq N_M^{[q]}$ for all $q = p^e$.
- (3) There exists $c \in R^\circ$ such that $cz^q\overline{\mathfrak{a}^q} \subseteq N_M^{[q]}$ for all $q = p^e$.

Proof. To prove (1) \Rightarrow (2), choose $d \in \mathfrak{a}^{r-1} \cap R^\circ$ and apply $d\tau(\mathfrak{a}^q) \subseteq \tau(\mathfrak{a}^{q+r-1}) \subseteq \mathfrak{a}^q$. As for (2) \Rightarrow (3), choose a test element $d \in \tau(R) \cap R^\circ$ and note that $d\overline{\mathfrak{a}^q} \subseteq \tau(\mathfrak{a}^q)$. □

Corollary 2.4. *Let (R, \mathfrak{m}) be a d -dimensional Noetherian local ring of characteristic $p > 0$ with infinite residue field. Then for any ideal $\mathfrak{a} \subseteq R$ and for any $n > 0$, one has $\tau(\mathfrak{a}^{n+d-1}) \subseteq \mathfrak{a}^n$. If, in addition, (R, \mathfrak{m}) is complete, then $\tau(\mathfrak{a}^{n+d-1}) \subseteq \tau(R)\mathfrak{a}^n$.*

Proof. First assume that \mathfrak{a} is an \mathfrak{m} -primary ideal. Then \mathfrak{a} has a minimal reduction $\mathfrak{q} \subseteq \mathfrak{a}$ generated by d elements, so that

$$\tau(\mathfrak{a}^{n+d-1}) = \tau(\mathfrak{q}^{n+d-1}) \subseteq \mathfrak{q}^n \subseteq \mathfrak{a}^n.$$

Next let \mathfrak{a} be an arbitrary ideal. Since our assertion holds true for every \mathfrak{m} -primary ideal, it follows that

$$\tau(\mathfrak{a}^{n+d-1}) \subseteq \bigcap_{t \in \mathbb{N}} \tau((\mathfrak{a} + \mathfrak{m}^t)^{n+d-1}) \subseteq \bigcap_{t \in \mathbb{N}} (\mathfrak{a} + \mathfrak{m}^t)^n = \mathfrak{a}^n.$$

The second assertion is proved in a similar way. □

Tight integral closure vs. \mathfrak{a} -tight closure. First, we recall the notion of tight integral closure, which was introduced by Hochster.

Definition 2.5 ([Ho2]). Let R be a Noetherian ring, and let $\{I_1, \dots, I_n\}$ be a set of ideals in R . An element $x \in R$ is in the *tight integral closure* $\{I_1, \dots, I_n\}^*$ if there exists $c \in R^\circ$ such that $cx^q \in \sum_{i=1}^n I_i^q$ for all sufficiently large $q = p^e$.

In [HWY1], the present authors have studied the F -rationality of Rees algebras $\mathbf{R}(I) = R[It]$ for \mathfrak{m} -primary ideals I , jointly with K.-i. Watanabe. One of the main results in [HWY1] is the following theorem, which gives a criterion for F -rationality of Rees algebras in terms of tight integral closure.

Theorem 2.6 (cf. [HWY1, Theorem 2.2]). *Let (R, \mathfrak{m}) be an excellent Cohen–Macaulay normal local ring of characteristic $p > 0$ with infinite residue field. Let I be an \mathfrak{m} -primary ideal of R and J its minimal reduction. Fix any system of parameters x_1, \dots, x_d for R generating J and put $J^{[l]} = (x_1^l, \dots, x_d^l)$ for $l \geq 1$. Then the Rees algebra $\mathbf{R}(I) = R[It]$ is F -rational if and only if $\mathbf{R}(I)$ is Cohen–Macaulay and the following equalities hold:*

$$\{I^{dl-r}, x_1^l R, \dots, x_d^l R\}^* = I^{dl-r} + J^{[l]} \text{ for all } l, r \geq 1 \text{ with } 1 \leq r \leq dl - 1.$$

We will show that all tight integral closures appearing in the above theorem can be represented as the form of some “ \mathfrak{a} -tight closure.” Namely, we have the following theorem.

Theorem 2.7. *Let (R, \mathfrak{m}) be an excellent equidimensional reduced local ring of characteristic $p > 0$ with $d = \dim R \geq 1$. Also, let x_1, \dots, x_d be a system of parameters of R , and put $J = (x_1, \dots, x_d)R$. Then we have*

$$\{J^{dl-r}, x_1^l R, \dots, x_d^l R\}^* = (x_1^l, \dots, x_d^l)^{*J^r}$$

for all integers $l, r \geq 1$.

Before proving the above theorem, we give some corollaries. We now recall that

$$\overline{I_1} + \dots + \overline{I_n} \subseteq \{I_1, \dots, I_n\}^*;$$

see [Ho1, Proposition 1.4].

Corollary 2.8. *Using the same notation as in Theorem 2.7, if $J \subseteq I \subseteq \overline{J}$, then $(J^{[l]})^{*J^r} \supseteq \overline{J^{dl-r}} + J^{[l]}$ for all $l, r \geq 1$. In particular, we have*

$$(1) \quad J + \overline{I^{d-1}} \subseteq J^*I.$$

(2) If $J^{*I^r} = J$, then $I^{d-r} \subseteq J$.

As an application of Theorem 2.7, we can rewrite Theorem 2.6 as follows.

Corollary 2.9. *Using the same notation as in Theorem 2.6, the Rees algebra $\mathbf{R}(I)$ is F -rational if and only if $\mathbf{R}(I)$ is Cohen–Macaulay and the following equalities hold:*

$$(J^{[l]})^{*I^r} = I^{dl-r} + J^{[l]} \text{ for all } l, r \geq 1 \text{ with } 1 \leq r \leq dl - 1.$$

Corollary 2.10. *Let R be an excellent F -rational local ring with $\dim R = 2$. Then for any parameter ideal J of R , we have $J^{*J} = \bar{J}$.*

Proof. Put $I = \bar{J}$. Then J is a minimal reduction of I . Since $\mathbf{R}(I)$ is F -rational by [HWY1, Theorem 3.1], it follows from Theorems 2.6 and 2.7 that $J^{*J} = J^{*I} = I + J = I$, as required. \square

In the following, we prove Theorem 2.7, and so we assume that (R, \mathfrak{m}, k) is an excellent equidimensional (not necessarily reduced) local ring of characteristic $p > 0$ with $d = \dim R \geq 1$. Also, let x_1, \dots, x_d be a system of parameters of R and put $J = (x_1, \dots, x_d)R$. Further, we set $J^{[l]} := (x_1^l, \dots, x_d^l)$ for all $l \geq 1$.

Lemma 2.11. *Suppose that x_1, \dots, x_d form a regular sequence. Then for all integers $l, r \geq 1$ we have*

$$(x_1^l, \dots, x_d^l) : (x_1, \dots, x_d)^r = (x_1, \dots, x_d)^{dl-r-d+1} + (x_1^l, \dots, x_d^l),$$

that is, $J^{[l]} : J^r = J^{dl-r-d+1} + J^{[l]}$.

Proof. The right-hand side is contained in the left-hand side because $J^{dl-r-d+1}J^r = J^{d(l-1)+1} \subseteq J^{[l]}$. We must show the opposite inclusion. To do that, let $w = x_1^{a_1} \cdots x_d^{a_d}$, where $0 \leq a_i \leq l-1$ for all i . First suppose that $\sum_{i=1}^d a_i = dl-r-d$. If we put $b_i = l-1-a_i$ for all i , then $b_i \geq 0$, $\sum_{i=1}^d b_i = r$ and $w \cdot x_1^{b_1} \cdots x_d^{b_d} = x_1^{l-1} \cdots x_d^{l-1} \notin J^{[l]}$. Next suppose that $\sum_{i=1}^d a_i \geq dl-r-d+1$. Then for all integers $c_i \geq 0$ with $\sum_{i=1}^d c_i = r$, we have $w \cdot x_1^{c_1} \cdots x_d^{c_d} = x_1^{a_1+c_1} \cdots x_d^{a_d+c_d} \in J^{[l]}$, because $\sum_{i=1}^d (a_i + c_i) \geq d(l-1) + 1$. Since $J^{[l]} : J^r$ is generated by monomials in x_1, \dots, x_d , the assertion follows from the above argument. \square

Using the colon-capturing property of tight closure, we obtain the following.

Corollary 2.12. *In the above notation, for all $l, r \geq 1$, we have*

$$J^{[l]} : J^r \subseteq (J^{dl-r-d+1} + J^{[l]})^*.$$

Proof. First suppose that R is complete and reduced. If we put $S = k[[x_1, \dots, x_d]]$, then S is a complete regular local domain and R is a finitely generated torsion-free S -module. Also, if we put $J_0 = (x_1, \dots, x_d)S$ and $J_0^{[l]} = (x_1^l, \dots, x_d^l)S$, then $J = J_0R$ and $J^{[l]} = J_0^{[l]}R$. Using the colon-capturing property of tight closure and the previous lemma, we get

$$J^{[l]} : J^r \subseteq \left((J_0^{[l]} : J_0^r)R \right)^* = \left((J_0^{dl-r-d+1} + J_0^{[l]})R \right)^* = \left(J^{dl-r-d+1} + J^{[l]} \right)^*.$$

Next we consider the general case. Fix $l, r \geq 1$ and put $K = J^{dl-r-d+1} + J^{[l]}$. Applying the above argument to $\widehat{R}_{\text{red}} = \widehat{\bar{R}}_{\text{red}}$, we have

$$J^{[l]} \widehat{R}_{\text{red}} : J^r \widehat{R}_{\text{red}} \subseteq (K \widehat{R}_{\text{red}})^* = (K \widehat{R})^* \widehat{R}_{\text{red}},$$

and hence $J^{[l]}\widehat{R} : J^r\widehat{R} \subseteq (K\widehat{R})^*$. By [BH, Proposition 10.3.18], we get

$$J^{[l]} : J^r = (J^{[l]}\widehat{R} : J^r\widehat{R}) \cap R = (K\widehat{R})^* \cap R = K^*\widehat{R} \cap R = K^*,$$

as required. \square

We are now ready to prove Theorem 2.7.

Proof of Theorem 2.7. Note that R admits a test element $c' \in R^\circ$, because R is an excellent reduced local ring ([HH2, Theorem 6.1]).

Let $z \in (J^{[l]})^{*J^r}$. By definition, there exists $c'' \in R^\circ$ such that $c''z^q J^{rq} \subseteq J^{[lq]}$ for all $q = p^e$, $e \gg 0$. Corollary 2.12 implies that

$$c''z^q \in J^{[lq]} : J^{rq} \subseteq (J^{(dl-r)q-d+1} + J^{[lq]})^*,$$

and hence

$$c'c''z^q \in J^{(dl-r)q-d+1} + J^{[lq]}$$

for all $q = p^e$, $e \gg 0$. Take any element $c''' \in J^{d-1} \cap R^\circ$ and put $c = c'c''c''' \in R^\circ$. Then $cz^q \in J^{(dl-r)q} + J^{[lq]}$ for all $q = p^e$, $e \gg 0$. Thus $z \in \{J^{dl-r}, x_1^l R, \dots, x_d^l R\}^*$.

Next we prove the opposite inclusion. Let $w \in \{J^{dl-r}, x_1^l R, \dots, x_d^l R\}^*$. By definition, there exists $c'' \in R^\circ$ such that $c''w^q \in J^{(dl-r)q} + J^{[lq]}$ for all $q = p^e$, $e \gg 0$. Thus $c''w^q J^{rq} \subseteq J^{dlq} + J^{[lq]}$.

On the other hand, by virtue of the tight closure Briançon–Skoda theorem [HH1, Theorem 5.4], we have

$$J^{dlq} \subseteq \overline{J^{dlq}} \subseteq (J^{[lq]})^*.$$

Taking a test element $c' \in R^\circ$, we have $c'(J^{[lq]})^* \subseteq J^{[lq]}$ for all $q = p^e$, $e \gg 0$. In particular, we have $c'c''w^q J^{rq} \subseteq J^{[lq]}$ for all sufficiently large $q = p^e$, and thus $w \in (J^{[l]})^{*J^r}$, as required. \square

Remark 2.13. Although we can easily see that $J^{*I}R_{\text{red}} \subseteq (JR_{\text{red}})^{*I}R_{\text{red}}$ always holds, we do not know whether or not the opposite inclusion holds. If it were true, we could remove the assumption that “ R is reduced” in Theorem 2.7.

A characterization of regular local rings. Let (R, \mathfrak{m}, k) be an excellent equidimensional reduced local ring of characteristic $p > 0$ with $d = \dim R \geq 1$. Then R is F -rational if and only if $J^* = J$ for some (every) parameter ideal J of R ; see [HH2] and [FW].

Let J be a minimal reduction of \mathfrak{m} . Since $J^* = J^{*R}$, we have the following increasing sequence of ideals in R :

$$J \subseteq J^* \subseteq J^{*\mathfrak{m}} \subseteq J^{*\mathfrak{m}^2} \subseteq \dots \subseteq J^{*\mathfrak{m}^{d-1}} \subseteq J^{*\mathfrak{m}^d} = R,$$

where the equality on the right follows from the tight closure Briançon–Skoda theorem ([HH1, Theorem 5.4]). So it is natural to ask the following question.

Question 2.14. Let R be an excellent equidimensional local ring of characteristic $p > 0$, and let J be a minimal reduction of \mathfrak{m} (in general, a parameter ideal of R). When does the equality $J^{*\mathfrak{m}^{d-1}} = J$ hold?

In the following, we give a characterization of regular local rings in terms of α -tight closure, which gives an answer to the above question. See also Section 5 about related problems.

Theorem 2.15. *Let (R, \mathfrak{m}, k) be an excellent equidimensional reduced local ring of characteristic $p > 0$ with $d = \dim R \geq 1$, and assume that k is infinite. Then the following conditions are equivalent:*

- (1) R is regular;
- (2) $\tau(\mathfrak{m}^{d-1}) = R$, i.e., $I^*\mathfrak{m}^{d-1} = I$ holds for every ideal I of R ;
- (3) $J^*\mathfrak{m}^{d-1} = J$ holds for some parameter ideal J of R .

In order to prove Theorem 2.15, we need the following lemma. Note that we do not need to assume that R is excellent in the proof of this lemma.

Lemma 2.16. *Let (R, \mathfrak{m}) be a regular local ring with $d = \dim R \geq 1$. Then $I^*\mathfrak{m}^{d-1} = I$ for every ideal I of R .*

Proof. Suppose that $I^*\mathfrak{m}^{d-1} \neq I$ for some ideal I of R . Let $z \in I^*\mathfrak{m}^{d-1} \setminus I$. By definition, there exists $c \in R^\circ$ such that $cz^q\mathfrak{m}^{(d-1)q} \subseteq I^{[q]}$ for all $q = p^e$, $e \gg 0$. Since the Frobenius map $F: R \rightarrow R$ is flat by Kunz' theorem ([Ku]), we have

$$c\mathfrak{m}^{(d-1)q} \subseteq I^{[q]} : z^q = (I : z)^{[q]} \subseteq \mathfrak{m}^{[q]},$$

and hence

$$c \in \mathfrak{m}^{[q]} : \mathfrak{m}^{(d-1)q} = \mathfrak{m}^{q-d+1}$$

for all $q = p^e$, $e \gg 0$, by Lemma 2.11. This is a contradiction. \square

Proof of Theorem 2.15. Note that R is approximately Gorenstein by our assumption. (1) \Rightarrow (2) follows from Lemma 2.16. Also, (2) \Rightarrow (3) is trivial.

Let us prove (3) \Rightarrow (1). Take a parameter ideal J such that $J^*\mathfrak{m}^{d-1} = J$. Since $J \subseteq J^* \subseteq J^*\mathfrak{m}^{d-1}$, we have $J = J^*$. Hence R is F-rational, and thus is Cohen–Macaulay. Then for a minimal reduction \mathfrak{q} of \mathfrak{m} , we have $\mathfrak{q}^*\mathfrak{m}^{d-1} = \mathfrak{q}$ (see 1.14). Thus we may assume that J is a minimal reduction of \mathfrak{m} . Then by virtue of Corollary 2.8, we have $\mathfrak{m} \subseteq J$. This implies that R is regular, as required. \square

3. INTERPRETATION OF MULTIPLIER IDEALS VIA \mathfrak{a} -TIGHT CLOSURE

We posed Theorem 2.1 as a prime characteristic analogue of Lipman's "modified Briançon–Skoda theorem" [Li]. The original form of "modified Briançon–Skoda" is stated in terms of what is called "adjoint ideals" by Lipman. Recently, this notion is reformulated in the theory of "multiplier ideals" from a different point of view and plays an important role in birational algebraic geometry; see [Ei], [La]. Actually, one can define multiplier ideals with "rational coefficients"; cf. Section 6.

Definition 3.1. Let Y be a normal \mathbb{Q} -Gorenstein variety over a field of characteristic zero and let $\mathfrak{a} \subset \mathcal{O}_Y$ be a nonzero ideal sheaf. Let $f: X \rightarrow Y$ be a log resolution of the ideal \mathfrak{a} , that is, a resolution of singularities of Y such that the ideal sheaf $\mathfrak{a}\mathcal{O}_X$ is invertible, say, $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ for an effective divisor Z on X , and that the union $\text{Exc}(f) \cup \text{Supp}(Z)$ of the f -exceptional locus and the support of Z is a simple normal crossing divisor. Given a rational number $t \geq 0$, the *multiplier ideal* $\mathcal{J}(t \cdot \mathfrak{a}) = \mathcal{J}(\mathfrak{a}^t)$ associated to t and \mathfrak{a} is defined to be the ideal sheaf

$$\mathcal{J}(t \cdot \mathfrak{a}) = f_*\mathcal{O}_X([K_{X/Y} - tZ])$$

in \mathcal{O}_Y , where the \mathbb{Q} -divisor $K_{X/Y} = K_X - f^*K_Y$ is the discrepancy of f . For $t = 1$, we just denote $\mathcal{J}(\mathfrak{a}) := \mathcal{J}(1 \cdot \mathfrak{a})$. This definition is independent of the choice of a log resolution $f: X \rightarrow Y$ of \mathfrak{a} .

Remark 3.2. (1) If $\mathfrak{b} = \mathfrak{a}^n$ for a nonnegative integer n , then $\mathcal{J}(t \cdot \mathfrak{b}) = \mathcal{J}(tn \cdot \mathfrak{a})$, and this justifies the notation $\mathcal{J}(\mathfrak{a}^t)$ with “formal exponent” t . Henceforth we prefer the exponential notation $\mathcal{J}(\mathfrak{a}^t)$ rather than $\mathcal{J}(t \cdot \mathfrak{a})$; cf. Section 6.

(2) Multiplier ideals $\mathcal{J}(\mathfrak{a})$ have properties similar to those of the ideals $\tau(\mathfrak{a})$; see Proposition 1.11. Namely, if $\mathfrak{b} \subseteq \mathfrak{a}$, then $\mathcal{J}(\mathfrak{b}) \subseteq \mathcal{J}(\mathfrak{a})$, and if \mathfrak{b} is a reduction of \mathfrak{a} , then the equality $\mathcal{J}(\mathfrak{b}) = \mathcal{J}(\mathfrak{a})$ holds; $\mathcal{J}(\mathfrak{a})\mathfrak{b} \subseteq \mathcal{J}(\mathfrak{a}\mathfrak{b})$ for any $\mathfrak{a}, \mathfrak{b}$, and if \mathfrak{b} is locally principal, then the equality $\mathcal{J}(\mathfrak{a})\mathfrak{b} = \mathcal{J}(\mathfrak{a}\mathfrak{b})$ holds; $\mathcal{J}(\mathfrak{a}) \neq 0$ as long as $\mathfrak{a} \neq 0$; and if Y has only log terminal singularities, then $\mathfrak{a} \subseteq \mathcal{J}(\mathfrak{a})$. Also, using vanishing theorems in characteristic zero, one can prove a “modified Briançon–Skoda theorem” ([Li, Theorem 1.4.1], [La]): If \mathfrak{a} is generated by r elements, then $\mathcal{J}(\mathfrak{a}^{n+r-1}) \subseteq \mathfrak{a}^n$ for every $n \geq 0$; cf. Theorem 2.1. Later in Section 4, we study more about similarity of the ideals $\tau(\mathfrak{a})$ and $\mathcal{J}(\mathfrak{a})$.

3.3. Reduction to prime characteristic. It is proved in [Ha2] and [Sm2] that the multiplier ideal $\mathcal{J}(R)$ of the unit ideal in a normal \mathbb{Q} -Gorenstein ring R essentially of finite type over a field of characteristic zero coincides, after reduction to characteristic $p \gg 0$, with the test ideal $\tau(R)$. We generalize this result in Theorem 3.4 below. To state the result, we have to begin with a ring R and an ideal \mathfrak{a} in characteristic zero, and reduce them to characteristic $p \gg 0$ together with a log resolution $f: X \rightarrow \operatorname{Spec} R$ of \mathfrak{a} .

Let R be an algebra essentially of finite type over a field k of characteristic zero, and let $\mathfrak{a} \subseteq R$ be an ideal. One can choose a finitely generated \mathbb{Z} -algebra A contained in k and a subalgebra R_A of R essentially of finite type over A such that the natural map $R_A \otimes_A k \rightarrow R$ is an isomorphism and $\mathfrak{a}_A = \mathfrak{a} \cap R_A$ generates the ideal \mathfrak{a} of R . For a maximal ideal μ of A , we consider the base change to its residue field $\kappa = \kappa(\mu)$ over A to get a prime characteristic ring $R_\kappa = R_A \otimes_A \kappa$ and an ideal $\mathfrak{a}_\kappa = \mathfrak{a}_A R_\kappa$. The data consisting of $\kappa = \kappa(\mu)$, R_κ , \mathfrak{a}_κ is considered to be a “prime characteristic model” of the original data in characteristic zero, and we refer to such $(\kappa, R_\kappa, \mathfrak{a}_\kappa)$ for maximal ideals μ in a suitable dense open subset of $\operatorname{Spec} A$ as “reduction to characteristic $p \gg 0$ ” of (k, R, \mathfrak{a}) . Furthermore, given a morphism of schemes essentially of finite type over k (and even more, a commutative diagram consisting of a finite collection thereof), e.g., a log resolution $f: X \rightarrow \operatorname{Spec} R$ of \mathfrak{a} , we can reduce this entire setup to characteristic $p \gg 0$. (See [Ha1], [Ha2], [HH3], [Sm1], [Sm2] for more details.) We use the phrase “in characteristic $p \gg 0$ ” when we speak of such a setup reduced from characteristic zero to characteristic $p \gg 0$.

The main result of this section is the following theorem, which ensures the correspondence of the ideal $\tau(\mathfrak{a})$ and the multiplier ideal $\mathcal{J}(\mathfrak{a})$. See Theorem 6.7 for a generalization of this theorem to the case of “rational coefficients”.

Theorem 3.4. *Let R be a normal \mathbb{Q} -Gorenstein local ring essentially of finite type over a field, and let \mathfrak{a} be a nonzero ideal. Assume that $\mathfrak{a} \subseteq R$ is reduced from characteristic zero to characteristic $p \gg 0$, together with a log resolution of singularities $f: X \rightarrow Y = \operatorname{Spec} R$ such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ is invertible. Then*

$$\tau(\mathfrak{a}) = H^0(X, \mathcal{O}_X(\lceil K_{X/Y} \rceil - Z)).$$

3.5. The remainder of this section is devoted to proving Theorem 3.4. Our strategy is to reduce to the case where the ring R is quasi-Gorenstein by passing to a canonical covering; see [T] for a direct proof which does not use a canonical covering. Let (R, \mathfrak{m}) be a normal \mathbb{Q} -Gorenstein local ring with a canonical module ω_R , and let r be the least positive integer such that the r th symbolic power $\omega_R^{(r)}$ of ω_R is

isomorphic to R . Given a fixed isomorphism $\omega_R^{(r)} \cong R$, one has a natural ring structure of $S = \bigoplus_{i=0}^{r-1} \omega_R^{(i)}$. This is a quasi-Gorenstein local ring (i.e., $\omega_S \cong S$) with the maximal ideal $\mathfrak{m} \oplus \bigoplus_{i=1}^{r-1} \omega_R^{(i)}$, and we call S a *canonical covering* of R .

The following two lemmas make it possible to reduce the proof of the theorem to the quasi-Gorenstein case.

Lemma 3.6. *Let (R, \mathfrak{m}) be a normal local ring of characteristic $p > 0$, S a canonical covering of R as above, and assume that r is not divisible by p . Then for any ideal $\mathfrak{a} \subseteq R$,*

$$\tau(\mathfrak{a}S) \cap R = \tau(\mathfrak{a}).$$

Proof. Identical to the case where $\mathfrak{a} = R$; see [Ha2, Section 2], and also [Sm2]. \square

The following lemma is also proved entirely as in the same way as the case where $\mathfrak{a} = R$ [Sm2, Proposition 3.2], but we include the proof for the sake of completeness.

Lemma 3.7 (cf. [Sm1]). *Let (R, \mathfrak{m}) be a normal local ring essentially of finite type over a field of characteristic zero, and let S be a canonical covering of R as above. Then, for any ideal $\mathfrak{a} \subseteq R$ and any rational number $t \geq 0$,*

$$\mathcal{J}((\mathfrak{a}S)^t) \cap R = \mathcal{J}(\mathfrak{a}^t).$$

Proof. Let $\pi: \operatorname{Spec} S \rightarrow \operatorname{Spec} R$ be the canonical covering, and let $f: X \rightarrow \operatorname{Spec} R$ and $g: Y \rightarrow \operatorname{Spec} S$ be log resolutions of \mathfrak{a} and $\mathfrak{a}S$, respectively, which make the following diagram commute:

$$\begin{array}{ccc} Y & \xrightarrow{g} & \operatorname{Spec} S \\ \tilde{\pi} \downarrow & & \downarrow \pi \\ X & \xrightarrow{f} & \operatorname{Spec} R \end{array}$$

Let E be the reduced divisor on X supported on $\operatorname{Exc}(f) \cup \operatorname{Supp}(Z)$ and let G be the reduced divisor with the same support as $\tilde{\pi}^*E$. Then the log ramification formula (see, e.g., [Ka]) tells us that

$$(3.7.1) \quad K_{Y/S} + G = \tilde{\pi}^*(K_{X/R} + E) + P$$

for some effective divisor P on Y such that $\operatorname{codim}(\tilde{\pi}(P), X) \geq 2$.

Now let $u \in \mathcal{J}(\mathfrak{a}^t) = H^0(X, \mathcal{O}_X(\lceil K_{X/R} - tZ \rceil))$, i.e., $\operatorname{div}_X(u) + \lceil K_{X/R} - tZ \rceil \geq 0$. Since $\operatorname{Supp}(K_{X/R} - tZ) \subseteq E$, one has $K_{X/R} - tZ + E \geq \lceil K_{X/R} - tZ \rceil$, so that

$$\operatorname{div}_X(u) + K_{X/R} + E - tZ \geq 0,$$

and this is a strict inequality for the coefficient in each irreducible component of E . Pulling this back by $\tilde{\pi}$ and applying (3.7.1) give an inequality

$$\operatorname{div}_Y(u) + K_{Y/S} + G - t\tilde{\pi}^*Z \geq 0,$$

which is a strict inequality for the coefficient in each irreducible component of G . Since G is reduced, it follows that $\operatorname{div}_Y(u) + \lceil K_{Y/S} - t\tilde{\pi}^*Z \rceil \geq 0$. Hence $u \in H^0(X, \mathcal{O}_Y(\lceil K_{Y/S} - t\tilde{\pi}^*Z \rceil)) = \mathcal{J}((\mathfrak{a}S)^t)$.

Conversely, let $u \in \mathcal{J}((\mathfrak{a}S)^t) \cap R$ and fix any prime divisor D on X . To prove $u \in \mathcal{J}(\mathfrak{a}^t)$, it is enough to show that $a := v_D(u) + \operatorname{coeff}_D(\lceil K_{X/R} - tZ \rceil)$ is non-negative. If D is not f -exceptional, this follows because $\tilde{\pi}$ is étale and finite at the generic point of D and $\operatorname{div}_Y(u) + \lceil K_{Y/S} - t\tilde{\pi}^*Z \rceil \geq 0$ by $u \in \mathcal{J}((\mathfrak{a}S)^t)$. Now let

D be f -exceptional, F any prime divisor on Y dominating D (this in particular implies that $F \not\subseteq \text{Supp}(P)$), and let $e := \text{coeff}_F(\tilde{\pi}^*D)$. Then

$$(3.7.2) \quad v_F(u) + \text{coeff}_F(\tilde{\pi}^*([K_{X/R} - tZ])) = ea.$$

On the other hand, since $\text{div}_Y(u) + [K_{Y/S} - t\tilde{\pi}^*Z] \geq 0$, it follows from (3.7.1) that $\text{div}_Y(u) + [\tilde{\pi}^*(K_{X/R} + E) - G + P - t\tilde{\pi}^*Z] \geq 0$. Since $F \not\subseteq \text{Supp}(P)$, this implies that

$$(3.7.3) \quad v_F(u) + \text{coeff}_F([\tilde{\pi}^*(K_{X/R} - tZ)]) \geq \text{coeff}_F(-\tilde{\pi}^*E + G) = -e + 1.$$

It follows from (3.7.2) and (3.7.3) that $ea \geq -e + 1$, so that $a \geq 0$, as required. \square

Remark. The proof of Lemma 3.7 works not only for canonical coverings but also under a weaker assumption that $R \hookrightarrow S$ is finite and étale in codimension 1. It is desirable that Lemma 3.6 also holds for finite extensions that are étale in codimension 1, and this issue has been recently settled by Takagi ([HT], [T]).

Proposition 3.8. *Let (R, \mathfrak{m}) be a d -dimensional normal local ring of characteristic $p > 0$, and let \mathfrak{a} be a nonzero ideal. Let $f: X \rightarrow \text{Spec } R$ be a proper birational morphism from a normal scheme X such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ is an invertible sheaf, and denote the closed fiber of f by $E = f^{-1}(\mathfrak{m})$. Then one has an inclusion*

$$\text{Ker} \left(H_{\mathfrak{m}}^d(R) \xrightarrow{\delta} H_E^d(\mathcal{O}_X(Z)) \right) \subseteq 0_{H_{\mathfrak{m}}^d(R)}^{\ast \mathfrak{a}},$$

where $\delta: H_{\mathfrak{m}}^d(R) \rightarrow H_E^d(\mathcal{O}_X(Z))$ is an edge map $H_{\mathfrak{m}}^d(R) \rightarrow H_E^d(\mathcal{O}_X)$ of the spectral sequence $H_{\mathfrak{m}}^i(R^j f_* \mathcal{O}_X) \Rightarrow H_E^{i+j}(\mathcal{O}_X)$ followed by the natural map $H_E^d(\mathcal{O}_X) \rightarrow H_E^d(\mathcal{O}_X(Z))$.

Proof. First note that, for any $q = p^e$ and any $c \in \mathfrak{a}^q \subseteq H^0(X, \mathcal{O}_X(-qZ))$, we have the following commutative diagram with exact rows:

$$\begin{array}{ccccccc} 0 & \longrightarrow & \text{Ker}(\delta) & \longrightarrow & H_{\mathfrak{m}}^d(R) & \longrightarrow & H_E^d(\mathcal{O}_X(Z)) \longrightarrow 0 \\ & & \downarrow & & \downarrow cF^e & & \downarrow cF^e \\ 0 & \longrightarrow & \text{Ker}(\delta) & \longrightarrow & H_{\mathfrak{m}}^d(R) & \longrightarrow & H_E^d(\mathcal{O}_X(Z)) \longrightarrow 0 \end{array}$$

Then $\mathfrak{a}^q F^e(\text{Ker}(\delta)) \subseteq \text{Ker}(\delta)$ for all $q = p^e$, and the conclusion follows from Proposition 1.15. \square

By virtue of Lemmas 3.6 and 3.7, it is sufficient to prove Theorem 3.4 in the case where R is quasi-Gorenstein, i.e., $\omega_R \cong R$. In this case, however, the assertion of Theorem 3.4 coincides with the following.

Theorem 3.9. *Let (R, \mathfrak{m}) be a d -dimensional normal local ring essentially of finite type over a field of characteristic p , and let \mathfrak{a} be a nonzero ideal. Assume that $\mathfrak{a} \subseteq R$ is reduced from characteristic zero to characteristic $p \gg 0$, together with a resolution of singularities $f: X \rightarrow Y = \text{Spec } R$ such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ is invertible. Then*

$$0_{H_{\mathfrak{m}}^d(R)}^{\ast \mathfrak{a}} = \text{Ker} \left(H_{\mathfrak{m}}^d(R) \xrightarrow{\delta} H_E^d(\mathcal{O}_X(Z)) \right),$$

where E is the closed fiber of f and δ is the edge map as in 3.8, or dually,

$$\text{Ann}_{\omega_R}(0_{H_{\mathfrak{m}}^d(R)}^{\ast \mathfrak{a}}) = H^0(X, \omega_X(-Z)) \quad \text{in } \omega_R.$$

Proof. First, let us discuss the situation in characteristic zero before reduction to characteristic $p \gg 0$. In characteristic zero, we choose a nonzero element $c \in \mathfrak{a}$ such that R_c is regular, and a log resolution $f: X \rightarrow \operatorname{Spec} R$ of the ideal $c\mathfrak{a}$. Then $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$ for an effective divisor Z on X , and $\operatorname{Supp}(Z + \operatorname{div}_X(c))$ is a simple normal crossing divisor. We choose an f -ample f -exceptional \mathbb{Q} -Cartier divisor D and a sufficiently small rational number $\varepsilon > 0$ such that $\lfloor \tilde{Z} + \varepsilon \operatorname{div}_X(c) \rfloor = Z$, where $\tilde{Z} = Z - D$.

Now we reduce the entire setup as above to characteristic $p \gg 0$, and switch the notation to denote things after reduction modulo p . Let the ideal \mathfrak{a} be generated by r elements. Since R_c is regular, some power c^s of c ($\in \mathfrak{a}$) is a usual test element and also an \mathfrak{a} -test element. Also, since $-\tilde{Z}$ is f -ample, $\mathcal{K} = \bigoplus_{n \geq 0} H^0(X, \omega_X(-[n\tilde{Z}]))$ is a finitely generated module over $\mathcal{R} = \bigoplus_{n \geq 0} H^0(X, \mathcal{O}_X(-n\tilde{Z}))$. Say \mathcal{K} is generated in degree $\leq n_0$.

Since we are working in characteristic $p \gg 0$, the e -times iterated Frobenius map

$$F^e: H^d_E(\mathcal{O}_X(\tilde{Z})) \rightarrow H^d_E(\mathcal{O}_X(p^e(\tilde{Z} + \varepsilon \operatorname{div}_X(c))))$$

is injective for all $q = p^e$ by Proposition 3.6 and Corollary 3.8 of [Ha1]; see also [Ha2, Discussion 4.6] and [MS]. This implies that the map

$$c^m F^e: H^d_E(\mathcal{O}_X(Z)) = H^d_E(\mathcal{O}_X(\tilde{Z})) \rightarrow H^d_E(\mathcal{O}_X(p^e \tilde{Z}))$$

is injective for all sufficiently large $e \in \mathbb{N}$ such that $p^e \varepsilon \geq m := r + 2s + n_0 - 1$. For such $q = p^e \gg 0$, we consider the following commutative diagram with exact rows:

$$\begin{array}{ccccccc} 0 & \longrightarrow & \operatorname{Ker}(\delta) & \longrightarrow & H^d_{\mathfrak{m}}(R) & \xrightarrow{\delta} & H^d_E(\mathcal{O}_X(Z)) \longrightarrow 0 \\ & & \downarrow & & \downarrow c^m F^e & & \downarrow c^m F^e \\ 0 & \longrightarrow & \operatorname{Ker}(\delta_e) & \longrightarrow & H^d_{\mathfrak{m}}(R) & \xrightarrow{\delta_e} & H^d_E(\mathcal{O}_X(q\tilde{Z})) \longrightarrow 0 \end{array}$$

Here, $\operatorname{Ker}(\delta_e)$ is considered to be the annihilator in $H^d_{\mathfrak{m}}(R)$ of $H^0(X, \omega_X(-[q\tilde{Z}]))$ viewed as a submodule of ω_R , with respect to the duality pairing $\omega_R \times H^d_{\mathfrak{m}}(R) \rightarrow H^d_{\mathfrak{m}}(\omega_R) \cong E_R(R/\mathfrak{m})$.

Now, if $\xi \in H^d_{\mathfrak{m}}(R)$ is not in $\operatorname{Ker}(\delta)$, then

$$c^m \xi^q \notin \operatorname{Ker}(\delta_e) = \operatorname{Ann}_{H^d_{\mathfrak{m}}(R)} H^0(X, \omega_X(-[q\tilde{Z}]))$$

for all $q = p^e \gg 0$ by the above commutative diagram. Then for all sufficiently large $q = p^e$ ($\geq n_0$), there exists an integer n with $0 \leq n \leq n_0$ such that

$$c^m \xi^q \notin [0 : H^0(X, \mathcal{O}_X((n - q)\tilde{Z}))]_{H^d_{\mathfrak{m}}(R)},$$

since \mathcal{K} is generated in degree $\leq n_0$ as a graded \mathcal{R} -module. Hence it follows from $H^0(X, \mathcal{O}_X((n - q)\tilde{Z})) \subseteq H^0(X, \mathcal{O}_X((n - q)Z)) \subseteq \overline{\mathfrak{a}^{q-n_0}}$ that $c^m \xi^q \overline{\mathfrak{a}^{q-n_0}} \neq 0$. On the other hand, since $c \in \mathfrak{a}$ and $c^s \in \tau(R)$,

$$c^{m-s} \overline{\mathfrak{a}^{q-n_0}} = c^{r+s+n_0-1} \overline{\mathfrak{a}^{q-n_0}} \subseteq \tau(R) \overline{\mathfrak{a}^{q+r-1}} \subseteq \mathfrak{a}^q$$

by the tight closure Briançon–Skoda theorem. Thus we have that $c^s \xi^q \mathfrak{a}^q \neq 0$. But this implies that $\xi \notin 0^{*\mathfrak{a}}_{H^d_{\mathfrak{m}}(R)}$, since c^s is an \mathfrak{a} -test element (cf. Theorem 1.13).

Consequently, we have $0^{*\mathfrak{a}}_{H^d_{\mathfrak{m}}(R)} \subseteq \operatorname{Ker}(\delta)$. The reverse inclusion follows from Proposition 3.8, and we are done. □

4. PROPERTIES OF THE IDEAL $\tau(\mathfrak{a})$ ANALOGOUS TO MULTIPLIER IDEALS

In this section we prove some properties of the ideal $\tau(\mathfrak{a})$ analogous to those of the multiplier ideal $\mathcal{J}(\mathfrak{a})$, which are found in [DEL], [How], and in Lazarsfeld's lecture notes [La]; see also similar results for "tight closure for pairs" in [T]. The results in this section are proved in any fixed characteristic $p > 0$. However, in view of Theorem 3.4, we can also say that they provide characteristic p proofs of the properties of multiplier ideals in characteristic zero.

We also want to remark that the results in this section can be generalized to "rational coefficients" (see Section 6), just like those for multiplier ideals.

Theorem 4.1 (Restriction theorem, cf. [La]). *Let (R, \mathfrak{m}) be a normal \mathbb{Q} -Gorenstein complete local ring of characteristic $p > 0$ and let $x \in \mathfrak{m}$ be a non-zero-divisor of R . Let $S = R/xR$ and assume that S is normal. Then for any ideal \mathfrak{a} of R ,*

$$\tau(\mathfrak{a}S) \subseteq \tau(\mathfrak{a})S.$$

Proof. Let $E_R = E_R(R/\mathfrak{m})$ and $E_S = E_S(S/\mathfrak{m}S)$ be the injective envelopes of residue fields of R and S , respectively. Then one has $E_S \cong (0 : x)_{E_R} \subset E_R$. We first prove the following claim, viewing E_S as a submodule of E_R via this inclusion.

Claim 4.1.1. $0_{E_R}^{*\mathfrak{a}} \cap E_S \subseteq 0_{E_S}^{*\mathfrak{a}S}.$

Proof of Claim 4.1.1. Since R is normal, we can choose an \mathfrak{a} -test element c whose image \bar{c} in S is nonzero by Proposition 1.7. Then we have the following commutative diagram for each $q = p^e$ (see the proof of [HW, Theorem 4.9]):

$$\begin{array}{ccc} E_S & \xrightarrow{\alpha} & E_R \\ \downarrow \bar{c}F_S^e & & \downarrow cx^{q-1}F_R^e \\ \mathbb{F}_S^e(E_S) & \xrightarrow{\beta} & \mathbb{F}_R^e(E_R) \end{array}$$

The map $\alpha: E_S \rightarrow E_R$ is the inclusion map mentioned above. Note also that $\mathbb{F}_R^e(E_R) \cong H_{\mathfrak{m}}^d(\omega_R^{(q)})$ and $\mathbb{F}_S^e(E_S) \cong H_{\mathfrak{m}}^{d-1}(\omega_S^{(q)})$ by [Wa], and $\beta: \mathbb{F}_S^e(E_S) \rightarrow \mathbb{F}_R^e(E_R)$ arises as a connecting homomorphism of the long exact sequence

$$\cdots \rightarrow H_{\mathfrak{m}}^{d-1}(\omega_R^{(q)}) \xrightarrow{x} H_{\mathfrak{m}}^{d-1}(\omega_R^{(q)}) \rightarrow H_{\mathfrak{m}}^{d-1}(\omega_S^{(q)}) \xrightarrow{\beta} H_{\mathfrak{m}}^d(\omega_R^{(q)}) \xrightarrow{x} H_{\mathfrak{m}}^d(\omega_R^{(q)}) \rightarrow 0$$

associated to $0 \rightarrow \omega_R^{(q)} \xrightarrow{x} \omega_R^{(q)} \rightarrow \omega_R^{(q)}/x\omega_R^{(q)} \rightarrow 0$ ([HW]). It then follows that $\text{Ker}(\beta) \cong H_{\mathfrak{m}}^{d-1}(\omega_R^{(q)})/xH_{\mathfrak{m}}^{d-1}(\omega_R^{(q)})$ is a proper S -submodule of $H_{\mathfrak{m}}^{d-1}(\omega_S^{(q)})$. Hence we can show as in the proof of Proposition 1.15 that $\text{Ker}(\beta)$ is annihilated by an element $\bar{d} \in S^\circ$.

Now let $\xi \in 0_{E_R}^{*\mathfrak{a}} \cap E_S$. Then $cF_R^e(\xi)\mathfrak{a}^q = 0$ in $\mathbb{F}_R^e(E_R)$ for all $q = p^e$, since c is an \mathfrak{a} -test element. This implies that $\bar{c}F_S^e(\xi)\mathfrak{a}^q \subseteq \text{Ker}(\beta)$ by the above commutative diagram. Therefore $\bar{c}\bar{d}F_S^e(\xi)(\mathfrak{a}S)^q = 0$ for all $q = p^e$ with $\bar{c}\bar{d} \in S^\circ$, whence $\xi \in 0_{E_S}^{*\mathfrak{a}S}$, as claimed. \square

We continue the proof of Theorem 4.1. Since R is complete, $0_{E_R}^{*\mathfrak{a}} = (0 : \tau(\mathfrak{a}))_{E_R}$, so that

$$0_{E_R}^{*\mathfrak{a}} \cap E_S = (0 : \tau(\mathfrak{a}) + xR)_{E_R} = \left(0 : \frac{\tau(\mathfrak{a}) + xR}{xR}\right)_{E_S} = (0 : \tau(\mathfrak{a})S)_{E_S}.$$

Hence by Claim 4.1.1, we conclude that $\tau(\mathfrak{a}S) \subseteq \text{Ann}_S(0_{E_R}^{*\mathfrak{a}} \cap E_S) = \tau(\mathfrak{a})S$. \square

Remark 4.2. Theorem 4.1 implies that F-regularity “deforms” in \mathbb{Q} -Gorenstein rings [AKM]. Namely, if R is \mathbb{Q} -Gorenstein, $x \in \mathfrak{m}$ is a non-zero-divisor and $S = R/xR$ is F-regular, then R is also F-regular. This result fails in the absence of \mathbb{Q} -Gorensteinness [Si]. On the other hand, the F-regularity of R does not imply the F-regularity of $S = R/xR$, and this suggests that the containment $\tau(\mathfrak{a}S) \subseteq \tau(\mathfrak{a})S$ is far from an equality in general; see also [HW, Theorem 4.9].

Our next objective is to apply Theorem 4.1 to show the property of $\tau(\mathfrak{a})$ called “subadditivity” (Theorem 4.5), which was established for multiplier ideals in regular rings by Demailly, Ein and Lazarsfeld [DEL]. Our strategy is to mimic the idea of “restriction to the diagonal” used in [DEL]. We first prove the following fact.

Lemma 4.3. *Let $R = k[[x_1, \dots, x_r]]$ and $S = k[[y_1, \dots, y_s]]$ be complete regular local rings with residue field k , and let $T = R \hat{\otimes}_k S = k[[x_1, \dots, x_r, y_1, \dots, y_s]]$ be their complete tensor product. Then for any ideals $\mathfrak{a} \subset R$ and $\mathfrak{b} \subset S$, we have that $(\mathfrak{a} \otimes S) \cap (R \otimes \mathfrak{b}) = (\mathfrak{a} \otimes \mathfrak{b})T$ in T .*

Proof. We regard R and S as subrings of T via the natural ring homomorphisms $R \hookrightarrow T$ and $S \hookrightarrow T$. Then what is to be proved is that $\mathfrak{a}T \cap \mathfrak{b}T = \mathfrak{a}\mathfrak{b}T$. To prove this we may assume, without loss of generality, that \mathfrak{b} is a proper ideal of S .

First we note that the composition $R \rightarrow T \rightarrow T/\mathfrak{b}T$ of the natural maps is a flat ring extension. Indeed, $R = R \otimes_k k \rightarrow R \otimes_k S/\mathfrak{b}$ is flat, since $k \rightarrow S/\mathfrak{b}$ is flat by $\mathfrak{b} \cap k = 0$, and the completion map $R \otimes_k S/\mathfrak{b} \rightarrow T/\mathfrak{b}T$ is also flat.

Now let F_\bullet be an R -free resolution of R/\mathfrak{a} . Then $F_\bullet \otimes_R T$ is a T -free resolution of $T/\mathfrak{a}T$, since T is flat over R . Since $T/\mathfrak{b}T$ is also flat over R , one has that $\mathrm{Tor}_i^T(T/\mathfrak{a}T, T/\mathfrak{b}T) = H_i((F_\bullet \otimes_R T) \otimes_T T/\mathfrak{b}) = H_i(F_\bullet \otimes_R T/\mathfrak{b}T) = 0$ for $i > 0$. In particular, $(\mathfrak{a}T \cap \mathfrak{b}T)/\mathfrak{a}\mathfrak{b}T \cong \mathrm{Tor}_1^T(T/\mathfrak{a}T, T/\mathfrak{b}T) = 0$. Thus we conclude that $\mathfrak{a}T \cap \mathfrak{b}T = \mathfrak{a}\mathfrak{b}T$, as required. \square

Proposition 4.4. *Let k be a field of characteristic p , and let $R = k[[x_1, \dots, x_r]]$, $S = k[[y_1, \dots, y_s]]$ and $T = R \hat{\otimes}_k S = k[[x_1, \dots, x_r, y_1, \dots, y_s]]$ be as in Lemma 4.3. Then, for any ideals $\mathfrak{a} \subset R$ and $\mathfrak{b} \subset S$,*

$$\tau((\mathfrak{a} \otimes \mathfrak{b})T) \subseteq (\tau(\mathfrak{a}) \otimes \tau(\mathfrak{b}))T.$$

Proof. Let us denote the injective envelopes of the residue fields of R , S , T by E_R , E_S , E_T , respectively. Then we can describe them in terms of inverse polynomials as $E_R = (x_1 \cdots x_r)^{-1}k[x_1^{-1}, \dots, x_r^{-1}]$, $E_S = (y_1 \cdots y_s)^{-1}k[y_1^{-1}, \dots, y_s^{-1}]$, $E_T = (x_1 \cdots x_r y_1 \cdots y_s)^{-1}k[x_1^{-1}, \dots, x_r^{-1}, y_1^{-1}, \dots, y_s^{-1}]$; so, in particular, $E_T = E_R \otimes_k E_S$. Then it is easy to see that $0_{E_R}^{*\mathfrak{a}} \otimes E_S + E_R \otimes 0_{E_S}^{*\mathfrak{b}} \subseteq 0_{E_T}^{*(\mathfrak{a} \otimes \mathfrak{b})T}$. Hence

$$\begin{aligned} \tau((\mathfrak{a} \otimes \mathfrak{b})T) &\subseteq \mathrm{Ann}_T(0_{E_R}^{*\mathfrak{a}} \otimes E_S) \cap \mathrm{Ann}_T(E_R \otimes 0_{E_S}^{*\mathfrak{b}}) \\ &= (\tau(\mathfrak{a}) \otimes S) \cap (R \otimes \tau(\mathfrak{b})) = (\tau(\mathfrak{a}) \otimes \tau(\mathfrak{b}))T \end{aligned}$$

by Lemma 4.3. \square

Theorem 4.5 (Subadditivity, cf. [DEL]). *Let (R, \mathfrak{m}) be a complete regular local ring of characteristic $p > 0$. Then for any two ideals $\mathfrak{a}, \mathfrak{b}$ of R ,*

$$\tau(\mathfrak{a}\mathfrak{b}) \subseteq \tau(\mathfrak{a})\tau(\mathfrak{b}).$$

Proof. Let $T = R \hat{\otimes}_k R$, and let $\Delta: T \rightarrow R$ be the ring homomorphism sending $x \otimes y \in T$ to $xy \in R$. If we restrict the containment $\tau(\mathfrak{a}\mathfrak{b}T) \subseteq \tau(\mathfrak{a})\tau(\mathfrak{b})T$ in

Proposition 4.4 by the diagonal map $\Delta: T \rightarrow R$, we immediately obtain

$$\tau(\mathfrak{a}\mathfrak{b}) \subseteq \tau(\mathfrak{a}\mathfrak{b}T)R \subseteq \tau(\mathfrak{a})\tau(\mathfrak{b})$$

by virtue of Theorem 4.1. □

Remark 4.6. If (R, \mathfrak{m}) is not F-regular, then it is clear that the subadditivity breaks down for $\mathfrak{a} = \mathfrak{b} = R$. We do not know whether or not the subadditivity always holds in F-regular rings in general.

4.7. Toric case. In [How], Howald gave a combinatorial description of the multiplier ideal $\mathcal{J}(\mathfrak{a})$ of a monomial ideal \mathfrak{a} in a polynomial ring over a field. We show that the ideal $\tau(\mathfrak{a})$ has a similar description in a more general situation, namely, \mathfrak{a} is a toric ideal of a toric ring R over a field k . Note that, in this case, the multiplier ideal $\mathcal{J}(\mathfrak{a})$ can be defined even if $\text{char } k = p > 0$, because there exists a log resolution of \mathfrak{a} in the toric category.

Let $M = \mathbb{Z}^d$, $N = \text{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$, and denote the duality pairing of $M_{\mathbb{R}} = M \otimes_{\mathbb{Z}} \mathbb{R}$ with $N_{\mathbb{R}} = N \otimes_{\mathbb{Z}} \mathbb{R}$ by $\langle \cdot, \cdot \rangle: M_{\mathbb{R}} \times N_{\mathbb{R}} \rightarrow \mathbb{R}$. Let $\sigma \subset N_{\mathbb{R}}$ be a strongly convex rational polyhedral cone, and denote $\sigma^{\vee} = \{m \in M_{\mathbb{R}} \mid \langle m, n \rangle \geq 0 \text{ for all } n \in \sigma\}$ as usual. Let $R = k[\sigma^{\vee} \cap M]$ be the toric ring over a field k defined by σ , that is, the subring of a polynomial ring $k[x_1, \dots, x_d]$ generated as a k -algebra by monomials $x^m = x_1^{m_1} \cdots x_d^{m_d}$ with $m = (m_1, \dots, m_d) \in \sigma^{\vee} \cap M$. Also, let D_1, \dots, D_s be the toric divisors of $\text{Spec } R$ corresponding to the primitive generators $n_1, \dots, n_s \in N$ of σ , respectively. A toric ideal $\mathfrak{a} \subseteq R$ is an ideal of R generated by monomials in x_1, \dots, x_d . Let $\mathfrak{a} \subseteq R$ be a toric ideal and let $P = P(\mathfrak{a}) \subset M_{\mathbb{R}}$ be the Newton polygon of \mathfrak{a} , that is, the convex hull of $\{m \in M \mid x^m \in \mathfrak{a}\}$ in $M_{\mathbb{R}}$. We denote the relative interior of P in $M_{\mathbb{R}}$ by $\text{Int}(P)$.

Now assume that R is \mathbb{Q} -Gorenstein. Then there exists $w \in M_{\mathbb{R}}$ such that $\langle w, n_i \rangle = 1$ for $i = 1, \dots, s$. Indeed, since $\omega_R^{(r)}$ is principally generated for some $r \in \mathbb{N}$ and $\omega_R^{(r)}$ corresponds to the divisor $-r \sum_{i=1}^s D_i$, we can write $\omega_R^{(r)} = x^{m_0} R$ for some $m_0 \in M$ such that $\langle m_0, n_i \rangle = v_{D_i}(x^{m_0}) = r$. Then set $w = m_0/r \in M_{\mathbb{R}}$.

Theorem 4.8. *Let $R = k[\sigma^{\vee} \cap M]$ be a \mathbb{Q} -Gorenstein toric ring over a field of characteristic $p > 0$, and let $w \in M_{\mathbb{R}}$ be as above. Then, for any toric ideal $\mathfrak{a} \subseteq R$,*

$$\tau(\mathfrak{a}) = \mathcal{J}(\mathfrak{a}),$$

and it is again a toric ideal. Moreover, for $m \in M$, the following conditions are equivalent to each other:

- (1) $x^m \in \tau(\mathfrak{a})$;
- (2) $m + w \in \text{Int}(P(\mathfrak{a}))$;
- (3) $x^m \in \mathcal{J}(\mathfrak{a})$.

Proof. We prove that $\tau(\mathfrak{a})$ is generated by monomials x^m satisfying the condition $m + w \in \text{Int}(P(\mathfrak{a}))$ in (2). It is essentially proved in [How] that $\mathcal{J}(\mathfrak{a})$ has the same property.

First, to simplify our argument, we note that $1 \in R^{\circ}$ is an \mathfrak{a} -test element, because toric rings are strongly F-regular. Hence, an element $z \in E$ of the injective envelope $E = E_R(R/\mathfrak{m})$ of the residue field $R/\mathfrak{m} = k$ is in $0_E^{*\mathfrak{a}}$ if and only if $z^q \mathfrak{a}^q = 0$ in $\mathbb{F}^e(E) = {}^eR \otimes_R E$ for all $q = p^e$.

Next we will compute the Frobenius map $F^e: E \rightarrow \mathbb{F}^e(E)$ explicitly. To do this we note that $\mathbb{F}^e(E) \cong H_{\mathfrak{m}}^d(\omega_R^{(q)})$ for $q = p^e$ by [Wa], and $H_{\mathfrak{m}}^d(\omega_R^{(q)})$ is k -dual to

$\omega_R^{(1-q)} = \bigoplus_{\langle m, n_i \rangle \geq 1-q} k \cdot x^m$. Therefore,

$$\mathbb{F}^e(E) = \bigoplus_{\langle m, n_i \rangle \leq q-1} k \cdot x^m = \bigoplus_{m \in (q-1)w - \sigma^\vee} k \cdot x^m,$$

and the Frobenius map $F^e: E \rightarrow \mathbb{F}^e(E)$ sends $x^m \in E$ to $x^{mq} \in \mathbb{F}^e(E)$.

It is now clear that $0_E^{*\mathfrak{a}}$ and hence $\tau(\mathfrak{a}) = \text{Ann}_R(0_E^{*\mathfrak{a}})$ are generated by monomials, because everything involved is \mathbb{Z}^s -graded.

We describe the \mathfrak{a} -tight closure $0_E^{*\mathfrak{a}}$ of zero in $E = \bigoplus_{u \in -\sigma^\vee \cap M} k \cdot x^u$. Let $u \in -\sigma^\vee \cap M$. Then $x^u \in 0_E^{*\mathfrak{a}}$ if and only if $x^{qu}\mathfrak{a}^q = 0$ in $\mathbb{F}^e(E)$ or, equivalently, $(qu+qP) \cap ((q-1)w - \sigma^\vee) \cap M = \emptyset$, for all $q = p^e$. Dividing out by q , we can rephrase this into the condition that $(u+P) \cap \text{Int}(w - \sigma^\vee) = \emptyset$, because $-w/q \in \text{Int}(-\sigma^\vee)$. Since this is equivalent to saying that $\text{Int}(u+P) \cap (w - \sigma^\vee) = \emptyset$, it follows that $x^u \in 0_E^{*\mathfrak{a}}$ if and only if $\text{Int}(P) \cap (w - u - \sigma^\vee) = \emptyset$ or, equivalently, if $w - u \notin \text{Int}(P)$.

Now the equivalence of conditions (1) and (2) follows immediately, because a monomial $x^m \in R = k[\sigma^\vee \cap M]$ is in $\tau(\mathfrak{a}) = \text{Ann}_R(0_E^{*\mathfrak{a}})$ if and only if $x^{-m} \notin 0_E^{*\mathfrak{a}}$. \square

Example 4.9. Let $S = k[x_1, \dots, x_d]$ be a polynomial ring and let $R = S^{(r)}$ be the r th Veronese subring of S . We can easily compute the ideal $\tau(\mathfrak{a}) = \mathcal{J}(\mathfrak{a})$ associated to a monomial ideal \mathfrak{a} of R as follows; cf. [How].

We choose N and M to be the overlattice $N = \mathbb{Z}^d + \frac{1}{r}(1, \dots, 1)\mathbb{Z}$ and the sublattice $M = \{m \in \mathbb{Z}^d \mid \langle m, n \rangle \in \mathbb{Z}\}$ of \mathbb{Z}^d , respectively, and let σ be the first orthant in $N_{\mathbb{R}} = \mathbb{R}^d$. Then the dual cone σ^\vee is also the first orthant in $M_{\mathbb{R}} = \mathbb{R}^d$, and the ring $R = k[\sigma^\vee \cap M]$ is realized as it is as the r th Veronese subring of $S = k[\sigma^\vee \cap \mathbb{Z}^d] = k[x_1, \dots, x_d]$. In this setting, the vector $w \in M_{\mathbb{R}}$ (for both R and S) defined in 4.7 is equal to $\mathbf{1} = (1, 1, \dots, 1)$. Also, for a monomial ideal $\mathfrak{a} \subseteq R$, the Newton polygons $P(\mathfrak{a})$ and $P(\mathfrak{a}S)$ are equal to each other in $M_{\mathbb{R}} = \mathbb{R}^d$. Therefore, Theorem 4.8 tells us that a monomial x^m in R (resp. S) is in $\tau(\mathfrak{a})$ (resp. $\tau(\mathfrak{a}S)$) if and only if $m + \mathbf{1} \in \text{Int}(P(\mathfrak{a})) = \text{Int}(P(\mathfrak{a}S))$, and, in particular,

$$\tau(\mathfrak{a}) = \tau(\mathfrak{a}S) \cap R;$$

cf. Lemma 3.6. For example, if $\mathfrak{a} = \mathfrak{m}^l$ is a power of the graded maximal ideal \mathfrak{m} of R , we have $\tau(\mathfrak{m}^l) = \mathfrak{m}^{\lceil l - (d-1)/r \rceil} = \mathfrak{m}^{\lfloor l + 1 - d/r \rfloor}$.

5. F-RATIONALITY OF REES ALGEBRAS AND THE BEHAVIOR OF $\tau(I)$

Throughout this section, we assume that (R, \mathfrak{m}) is an excellent Gorenstein local domain of characteristic $p > 0$, and that I is an \mathfrak{m} -primary ideal of R . Put $d = \dim R \geq 2$. Let $\mathbf{R}(I) = R[It]$ denote the Rees algebra of I over R , and $\mathfrak{M} = \mathfrak{m}\mathbf{R}(I) + \mathbf{R}(I)_+$, the unique homogeneous maximal ideal of $\mathbf{R}(I)$. We will denote by $\mathbf{R}'(I) = R[It, t^{-1}]$ and $G(I) = \mathbf{R}'(I)/t^{-1}\mathbf{R}'(I) = \bigoplus_{n \geq 0} I^n/I^{n+1}$ the extended Rees algebra and the associated graded ring of I , respectively. Also, let $\omega_{\mathbf{R}(I)}$ denote the graded canonical module of $\mathbf{R}(I)$, and let $\pi: Y = \text{Proj } \mathbf{R}(I) \rightarrow \text{Spec } R$ be the blowing-up with respect to I .

The main purpose of this section is to describe $\omega_{\mathbf{R}(I)}$ in terms of $\tau(I^n)$ under the assumption that $\mathbf{R}(I)$ is F-rational. Actually, we prove the following theorem.

Theorem 5.1. *Let (R, \mathfrak{m}) be an excellent Gorenstein local domain of characteristic $p > 0$ with $d = \dim R \geq 2$. Let I be an \mathfrak{m} -primary ideal of R and J its minimal*

reduction. Then $\tau(I) \subseteq J : I^{d-1}$. If we assume that $\mathbf{R}(I)$ is F -rational, then $\tau(I) = J : I^{d-1}$ and

$$\omega_{\mathbf{R}(I)} = \bigoplus_{n \geq 1} H^0(Y, I^n \omega_Y) \cong \bigoplus_{n \geq 1} \tau(I^n).$$

Discussion 5.2. The above theorem is motivated by Hyry's papers [Hy1], [Hy2]. For example, the description of $\omega_{\mathbf{R}(I)}$ in Theorem 5.1 corresponds to the following fact used in [Hy1]: Let (R, \mathfrak{m}) be a regular local ring essentially of finite type over a field of characteristic zero, and let I be an ideal of R . Suppose that $\text{Proj } \mathbf{R}(I)$ has rational singularities. The graded canonical module of $\mathbf{R}(I)$ is then $\omega_{\mathbf{R}(I)} = \bigoplus_{n \geq 1} \mathcal{J}(I^n)$.

Actually, if $\mathbf{R}(I)$ is F -rational, then so is $Y = \text{Proj } \mathbf{R}(I)$, whence Y is pseudo-rational [Sm1]. Therefore, if Y has a resolution of singularities $f: X \rightarrow Y$, then $H^0(X, I^n \omega_X) \cong H^0(Y, I^n \omega_Y)$ for every $n \geq 0$. The left-hand side of this equality coincides with the multiplier ideal $\mathcal{J}(I^n)$ via the isomorphism $\omega_R \cong R$ as long as $\mathcal{J}(I^n)$ is defined. Moreover, $[\omega_{\mathbf{R}(I)}]_n \cong H^0(Y, I^n \omega_Y)$, since $\mathbf{R}(I)$ is Cohen-Macaulay; see e.g. [HHK]. In particular, we have $\tau(I) = \mathcal{J}(I)$ in this case.

Consequently, Theorem 5.1 claims that the F -rationality of $\mathbf{R}(I)$ gives a sufficient condition for $\tau(I) = \mathcal{J}(I)$ to hold in any fixed positive characteristic. \square

We obtain the following corollary from Theorem 5.1 and [HWY1, Corollary 3.3].

Corollary 5.3. *Suppose that (R, \mathfrak{m}) is a two-dimensional rational double point. Let I be an \mathfrak{m} -primary integrally closed ideal of R , and J its minimal reduction. Then $\tau(I) \subseteq J : I (= \mathcal{J}(I))$. Also, $\mathbf{R}(I)$ is F -rational if and only if $\tau(I) = J : I$.*

One can easily check the following example using the method developed in [HWY1, Section 3].

Example 5.4 (cf. [HWY1, Theorem 3.1]). (1) Let (R, \mathfrak{m}) be a two-dimensional excellent Gorenstein F -rational local ring (i.e., F -rational double point), and I an \mathfrak{m} -primary integrally closed ideal of R . Then $\tau(I) = J : I$ for any minimal reduction J of I .

(2) Let $R = k[[x, y, z]]/(x^2 + y^3 + z^5)$, where k is an algebraically closed field of characteristic 2. Put $\mathfrak{m} = (x, y, z)R$ and $J = (y, z)R$. Then R is a two-dimensional rational double point, but not F -rational. Also, we have:

- (a) $R(\mathfrak{m})$ is not F -rational.
- (b) $J^{*\mathfrak{m}} = \mathfrak{m}$. In particular, $J : J^{*\mathfrak{m}} = \mathfrak{m}$.
- (c) $(J^{[2]})^{*\mathfrak{m}} = (y^2, z^2, xy)$. In particular, $J^{[2]} : (J^{[2]})^{*\mathfrak{m}} = (x, y, z^2)$.
- (d) $\tau(\mathfrak{m}) \subseteq (x, y, z^2) \subsetneq \mathcal{J}(\mathfrak{m}) = \mathfrak{m}$.

In the following, we will prove Theorem 5.1.

Lemma 5.5. *Let (R, \mathfrak{m}) be a Gorenstein local ring of any characteristic. Also, let I be an \mathfrak{m} -primary ideal of R and put $G(I) = \bigoplus_{n \geq 0} I^n/I^{n+1}$, the associated graded ring of I . Assume that $[H^d_{\mathfrak{M}}(\mathbf{R}(I))]_0 = [H^d_{\mathfrak{M}}(\mathbf{R}(I))]_{-1} = 0$ (e.g., $\mathbf{R}(I)$ is Cohen-Macaulay). Then $R/[\omega_{\mathbf{R}(I)}]_1 \cong ([H^d_{\mathfrak{M}}(G(I))]_{-1})^\vee$, where $(\)^\vee$ denotes the Matlis dual of R .*

Proof. Consider the following two standard exact sequences:

$$0 \longrightarrow \mathbf{R}(I)_+ \longrightarrow \mathbf{R}(I) \longrightarrow R \longrightarrow 0,$$

$$0 \longrightarrow \mathbf{R}(I)_+(1) \longrightarrow \mathbf{R}(I) \longrightarrow G(I) \longrightarrow 0.$$

From the first exact sequence, we have

$$0 = [H_{\mathfrak{m}}^d(\mathbf{R}(I))]_0 \rightarrow H_{\mathfrak{m}}^d(R) \rightarrow [H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I)_+)]_0 \rightarrow [H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I))]_0 = 0,$$

where the vanishing on the right follows from $a(\mathbf{R}(I)) = -1$. Since R is Gorenstein, $R = \omega_R \cong (H_{\mathfrak{m}}^d(R))^{\vee} \cong ([H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I)_+)]_0)^{\vee}$. On the other hand, the second exact sequence gives

$$0 = [H_{\mathfrak{m}}^d(\mathbf{R}(I))]_{-1} \rightarrow [H_{\mathfrak{m}}^d(G(I))]_{-1} \rightarrow [H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I)_+)]_0 \rightarrow [H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I))]_{-1} \rightarrow 0.$$

Dualizing the above sequence, we get

$$0 \longrightarrow [\omega_{\mathbf{R}(I)}]_1 \longrightarrow ([H_{\mathfrak{m}}^{d+1}(\mathbf{R}(I)_+)]_0)^{\vee} \cong R \longrightarrow ([H_{\mathfrak{m}}^d(G(I))]_{-1})^{\vee} \longrightarrow 0.$$

This yields the required assertion. □

Proposition 5.6. *Using the same notation as in Lemma 5.5, assume further that $\mathbf{R}(I)$ is Cohen–Macaulay. Then $[\omega_{\mathbf{R}(I)}]_1 = J : I^{d-1}$ for every minimal reduction J of I .*

Proof. Let x_1, x_2, \dots, x_d be a minimal system of generators of J . Put $G := G(I)$. Then it is well known that $G(I)$ is Cohen–Macaulay and the images in G of $x_1t, \dots, x_dt \in \mathbf{R}(I)_1$ form a regular sequence ([GS]). Setting $x_i^* := x_it \bmod I^2$ for each i , we have an exact sequence

$$0 \longrightarrow G(-1) \xrightarrow{x_1^*} G \longrightarrow G/x_1^*G \cong G(I/x_1R) \longrightarrow 0$$

by [VV]. From this sequence, we get an exact sequence

$$0 = H_{\mathfrak{m}}^{d-1}(G) \longrightarrow H_{\mathfrak{m}}^{d-1}(G/x_1^*G) \longrightarrow H_{\mathfrak{m}}^d(G)(-1) \xrightarrow{x_1^*} H_{\mathfrak{m}}^d(G) \longrightarrow 0.$$

Since $a(G) \leq -1$ ([GS]), we have $[H_{\mathfrak{m}}^{d-1}(G/x_1^*G)]_0 \cong [H_{\mathfrak{m}}^d(G)]_{-1}$ and $a(G/x_1^*G) \leq -1 + 1 = 0$. By repeating the above argument, we get

$$[H_{\mathfrak{m}}^d(G)]_{-1} \cong [H_{\mathfrak{m}}^0(G/(x_1^*, \dots, x_d^*)G)]_{d-1} \cong [H_{\mathfrak{m}}^0(G(I/J))]_{d-1} = \frac{J + I^{d-1}}{J}.$$

Also, since R/J is Gorenstein, we have $\left(\frac{J + I^{d-1}}{J}\right)^{\vee} \cong R/J : I^{d-1}$. Combining this with the previous lemma, we get

$$[\omega_{\mathbf{R}(I)}]_1 = \text{Ann}_R([H_{\mathfrak{m}}^d(G)]_{-1})^{\vee} = \text{Ann}_R\left(\frac{J + I^{d-1}}{J}\right)^{\vee} = J : I^{d-1}.$$

□

Proposition 5.7. *Let (R, \mathfrak{m}) be an excellent Gorenstein local domain of characteristic $p > 0$, and let I be an \mathfrak{m} -primary ideal of R . Also, let J be a minimal reduction of I . Then we have the following statements.*

- (1) $\tau(I) \subseteq J : J^*I \subseteq J : I^{d-1}$.
- (2) If $\mathbf{R}(I)$ is F -rational, then $\tau(I) = J : J^*I = J : I^{d-1}$.

Proof. Let x_1, \dots, x_d be a system of generators of J and put $J^{[l]} := (x_1^l, \dots, x_d^l)$ for all integers $l \geq 1$.

(1) One has $\tau(I) \subseteq J : J^*I \subseteq J : (J + I^{d-1}) = J : I^{d-1}$ by the definition of $\tau(I)$ and Corollary 2.8.

To see (2), we may assume that $\mathbf{R}(I)$ is Cohen–Macaulay. Then $I^d = JI^{d-1}$ ([GS]). Hence $I^{dl-1} = J^{dl-d}I^{d-1} = (J^{[l]}J^{dl-d-l} + (x_1 \cdots x_d)^{l-1}R)I^{d-1}$. Thus

$J^{[l]} + I^{dl-1} = J^{[l]} + (x_1 \cdots x_d)^{l-1} I^{d-1}$. In particular, for all $l \geq 1$, we have $J^{[l]} : (J^{[l]} + I^{dl-1}) = (J^{[l]} : (x_1 \cdots x_d)^{l-1}) : I^{d-1} = J : I^{d-1}$.

Now suppose that $\mathbf{R}(I)$ is F-rational. Then since $(J^{[l]})^{*I} = J^{[l]} + I^{dl-1}$ by Corollary 2.9, we have that $J^{[l]} : (J^{[l]})^{*I} = J : I^{d-1}$ for all $l \geq 1$. Hence $\tau(I) = J : J^{*I} = J : I^{d-1}$, as required. \square

Proof of Theorem 5.1. Note that $\mathbf{R}(I^n)$ is F-rational if $\mathbf{R}(I)$ is. Actually, it is a module-finite pure subring of $\mathbf{R}(I)$. Thus the required assertion immediately follows from Propositions 5.6 and 5.7. \square

Proof of Corollary 5.3. Let I be an \mathfrak{m} -primary integrally closed ideal and J its minimal reduction. Then it is well known that $I^2 = JI$, and thus $\mathbf{R}(I)$ is Cohen–Macaulay.

It is enough to show that $\tau(I) = J : I$ implies that $\mathbf{R}(I)$ is F-rational. Suppose that $\tau(I) = J : I$. Since $\tau(I) \subseteq J^{[l]} : (J^{[l]})^{*I} \subseteq J^{[l]} : (J^{[l]} + I^{2l-1}) = J : I$, in general, we have $J^{[l]} : (J^{[l]})^{*I} = J^{[l]} : (J^{[l]} + I^{2l-1})$. This implies that $(J^{[l]})^{*I} = J^{[l]} + I^{2l-1}$ for all $l \geq 1$, because $R/J^{[l]}$ is an Artinian Gorenstein local ring. By [HWY1, Corollary 3.3(2)], we conclude that $\mathbf{R}(I)$ is F-rational. \square

In the rest of this section, we will give some applications of Theorem 5.1. Before stating our results, let us recall the notion of a -invariant. Let I be an \mathfrak{m} -primary ideal of R , and put $G := G(I)$ and $\mathfrak{M} := \mathfrak{m}\mathbf{R}(I) + \mathbf{R}(I)_+$. Then the a -invariant $a(G)$ of G is defined by $a(G) := \max\{n \in \mathbb{Z} \mid [H_{\mathfrak{M}}^d(G)]_n \neq 0\}$. See [GW] for details.

Proposition 5.8. *Let (R, \mathfrak{m}) be an excellent Gorenstein local domain of characteristic $p > 0$. Let I be an \mathfrak{m} -primary ideal of R . Suppose that $\mathbf{R}(I)$ is F-rational and $G := G(I)$ is Gorenstein. Then $\tau(I^n) = I^{n+a(G)+1}$ for all integers $n \geq 1$.*

Proof. The F-rationality of $\mathbf{R}(I)$ implies that $\tau(I^n) = [\omega_{\mathbf{R}(I)}]_n = H^0(Y, I^n \omega_Y)$ for all $n \geq 1$; see Discussion 5.2. Also, since G is Cohen–Macaulay, we have

$$(5.8.1) \quad \omega_G \cong \bigoplus_{n \geq 1} H^0(Y, I^{n-1} \omega_Y) / H^0(Y, I^n \omega_Y)$$

and $R = H^0(Y, \omega_Y) = \cdots = H^0(Y, I^{-a-1} \omega_Y)$, where $a = a(G) \leq -1$; see e.g. [Hy2, Theorem 2.2]. On the other hand, since G is Gorenstein, we have

$$(5.8.2) \quad \omega_G \cong G(a) = \bigoplus_{n \geq -a} I^{n+a} / I^{n+a+1}.$$

Comparing (5.8.1) with (5.8.2), one can easily see that $\tau(I^n) = I^{n+a+1}$ by induction on $n \geq 1$. \square

Corollary 5.9. *Let (R, \mathfrak{m}) be an (excellent) regular local ring of characteristic $p > 0$. Then $\tau(\mathfrak{m}^n) = \mathfrak{m}^{n-d+1}$ for all $n \geq 1$.*

Proof. Suppose that R is a regular local ring. Then $R(\mathfrak{m})$ is F-rational and $G(\mathfrak{m}) \cong k[X_1, \dots, X_d]$ is Gorenstein with $a(G(\mathfrak{m})) = -d$. Hence we can apply the above proposition. \square

Remark 5.10. Corollary 5.9 is a generalization of the implication (1) \Rightarrow (2) in Theorem 2.15. This also follows from Theorem 4.8.

Let $J \subseteq I$ be ideals of R . Recall that the *coefficient ideal* of I relative to J , denoted by $\mathfrak{a}(I, J)$, is defined to be the largest ideal \mathfrak{a} of R for which $I\mathfrak{a} = J\mathfrak{a}$.

Remark 5.11. In [Hy2], Hyry proved that if R is a Gorenstein local ring and $\mathbf{R}(I)$ has rational singularities, then $\mathcal{J}(I^{d-1}) = \mathfrak{a}(I, J)$. In fact, a similar result follows from Theorem 5.1 and [Hy2, Theorem 3.4]: Let (R, \mathfrak{m}) be an excellent Gorenstein local domain of characteristic $p > 0$. Let I be an \mathfrak{m} -primary ideal of R , and J its minimal reduction. If $\mathbf{R}(I)$ is F-rational, then $\tau(I^{d-1}) = H^0(Y, I^{d-1}\omega_Y) = \mathfrak{a}(I, J)$. In particular, if, in addition, $I^2 = JI$, then $\tau(I^{d-1}) = J : I$.

In the rest of this section, we direct our attention to the ideal $\tau(\mathfrak{m})$. Let (R, \mathfrak{m}) be an excellent Gorenstein F-rational local domain of characteristic $p > 0$. Then $\tau(\mathfrak{m}) \supseteq \mathfrak{m}$, that is, $\tau(\mathfrak{m}) = \mathfrak{m}$ or $\tau(\mathfrak{m}) = R$. For example, if R is a regular local ring with $\dim R \geq 2$, then $\tau(\mathfrak{m}) = R$. More generally, we have the following proposition.

Proposition 5.12. *Let (R, \mathfrak{m}) be an excellent Gorenstein local domain of characteristic $p > 0$ with $d = \dim R \geq 2$. Suppose that there exists an \mathfrak{m} -primary ideal I such that $\mathbf{R}(I)$ is F-rational with $a(G(I)) \neq -1$. Then R is F-rational with $\tau(\mathfrak{m}) = R$.*

Proof. By virtue of [HWY1, Corollary 2.13], R is F-rational.

By Theorem 5.1, we have $\tau(I) = J : I^{d-1}$ for any minimal reduction J of I . Since $\mathbf{R}(I)$ is Cohen–Macaulay with $a(G(I)) \neq -1$, we have that $I^{d-1} = JI^{d-2} \subseteq J$. Hence $\tau(I) = R$. In particular, $\tau(\mathfrak{m}) = R$, because $\tau(\mathfrak{m}) \supseteq \tau(I)$. \square

In view of the above proposition it is natural to ask the following question.

Question 5.13. Let (R, \mathfrak{m}) be an excellent Gorenstein F-rational local domain of characteristic $p > 0$ with $\tau(\mathfrak{m}) = R$ and $\dim R \geq 2$. When is $\mathbf{R}(\mathfrak{m})$ F-rational then?

In the case of two-dimensional local rings, $\tau(\mathfrak{m}) = R$ implies that R is regular. Then $R(\mathfrak{m})$ is Gorenstein and F-rational with $a(G(\mathfrak{m})) = -2$. As for three-dimensional local rings, we have the following answer to the above question.

Proposition 5.14. *Let (R, \mathfrak{m}) be a three-dimensional excellent Gorenstein local ring that is not regular. Then the following conditions are equivalent.*

- (1) $\tau(\mathfrak{m}) = R$, that is, $I^*\mathfrak{m} = I$ holds for every ideal I of R .
- (2) $J^*\mathfrak{m} = J$ holds for some parameter ideal J of R .
- (3) $\mathbf{R}'(\mathfrak{m})$ is F-rational, and $\mathfrak{m}^2 = J\mathfrak{m}$ for some minimal reduction J of \mathfrak{m} .
- (4) $\mathbf{R}(\mathfrak{m})$ is Gorenstein and F-rational.
- (5) $\tau(\mathfrak{m}^n) = \mathfrak{m}^{n-1}$ holds for all integers $n \geq 1$.

Proof. (1) \Rightarrow (2) and (5) \Rightarrow (1) are trivial. (4) \Rightarrow (5) follows from Proposition 5.8, since $a(G(\mathfrak{m})) = -2$ ([GS]).

To see that (2) \Rightarrow (3), we may assume that J is a minimal reduction of \mathfrak{m} ; see Discussion 1.14. By Corollary 2.8, we have $\mathfrak{m}^2 \subseteq J$, and thus $\mathfrak{m}^2 = J\mathfrak{m}$. Also, $\mathbf{R}'(\mathfrak{m})$ is F-rational by [HWY1, Corollary 4.5].

(3) \Rightarrow (4): Note that a Gorenstein local ring having minimal multiplicity is a hypersurface with multiplicity at most 2. Thus R and $G(\mathfrak{m})$ are hypersurfaces, and $a(G(\mathfrak{m})) = 1 - \dim R = -2$. Hence $\mathbf{R}(\mathfrak{m})$ is Gorenstein ([GS]). Also, since $\mathbf{R}'(\mathfrak{m})$ is Gorenstein and F-rational, $\mathbf{R}(\mathfrak{m})$ is F-regular. \square

Discussion 5.15. We can generalize the equivalence of (2) and (3) in Proposition 5.14 as follows; see also Theorem 2.15.

Let (R, \mathfrak{m}) be an excellent equidimensional reduced local ring of characteristic $p > 0$. Then the following conditions are equivalent.

- (1) $\mathbf{R}'(\mathfrak{m})$ is F -rational, and $\mathfrak{m}^2 = J\mathfrak{m}$ for some minimal reduction J of \mathfrak{m} .
- (2) $J^*\mathfrak{m}^{d-2} = J$ holds for every (or equivalently, some) parameter ideal J of A .

If, in addition, R is Gorenstein, then the following condition is also equivalent to the above conditions:

- (3) $\tau(\mathfrak{m}^{d-2}) = R$.

□

Example 5.16 (cf. [HWY2, Proposition 3.12], [HWY1, Sect. 5]). Let (R, \mathfrak{m}) be an excellent three-dimensional Gorenstein normal local domain of characteristic $p > 0$. If R admits a non-zero-divisor $f \in \mathfrak{m}$ such that R/fR is F -rational, then $\tau(\mathfrak{m}) = R$.

For example, let $R = k[[x, y, z, w]]/(x^2 + y^a + z^b + w^c)$, where k is a field of characteristic $p > 0$, and a, b, c are integers with $2 \leq a \leq b \leq c \ll p$. If $1/2 + 1/a + 1/b > 1$, then $\tau(\mathfrak{m}) = R$. Otherwise, $\tau(\mathfrak{m}) = \mathfrak{m}$.

Remark 5.17. (1) If R is a three-dimensional regular local ring, then $\tau(\mathfrak{m}^2) = R$ (and thus $\tau(\mathfrak{m}) = R$) and $\mathbf{R}(\mathfrak{m})$ is F -rational. But $\mathbf{R}(\mathfrak{m})$ is not Gorenstein, and $\tau(\mathfrak{m}^n) = \mathfrak{m}^{n-2}$ for all $n \geq 1$.

(2) We have no examples of a Gorenstein local ring R for which $\tau(\mathfrak{m}) = R$ but $\mathbf{R}(\mathfrak{m})$ is not F -rational.

Discussion 5.18. Let (R, \mathfrak{m}) be a complete regular local ring of characteristic $p > 0$ with $d = \dim R \geq 2$, and let I be an \mathfrak{m} -primary ideal of R . Then we expect that $\tau(I) \supsetneq I$.

For example, this is true if $R(I)$ is F -rational. We sketch a proof here. Suppose that $\tau(I) = I$. Then $\tau(I^n) \subseteq \tau(I)^n = I^n$ for all $n \geq 1$, by the subadditivity (Theorem 4.5). On the other hand, since R is F -regular, we have $\tau(I^n) \supseteq I^n$ in general. Also, by Theorem 5.1, we have $\omega_{\mathbf{R}(I)} = \bigoplus_{n \geq 1} I^n = \mathbf{R}(I)_+$. In particular, since $\mathbf{R}(I)/\omega_{\mathbf{R}(I)} \cong R$ is regular, so is $\mathbf{R}(I)$. (Note: Recently, S. Goto et. al. proved a more general result.) But this is impossible, because $\dim R \geq 2$. Hence $\tau(I) \supsetneq I$, as required.

As for multiplier ideals, the authors are informed of the following result by K.-i. Watanabe: Let R be a regular local ring essentially of finite type over a field of characteristic zero. Then $\mathcal{J}(I) \supsetneq I$ for any \mathfrak{m} -primary ideal I of R . □

6. RATIONAL COEFFICIENTS

Recently, the theory of multiplier ideals with “rational coefficients” has been developed and applied successfully to various problems in algebraic geometry and commutative algebra ([ELS], [La]). This motivates us to extend the notions of \mathfrak{a} -tight closure and the ideal $\tau(\mathfrak{a})$ to those with “rational coefficients”. In this last section we make a few remarks on rational coefficients and address the results that generalize in this setting.

Definition 6.1. Let \mathfrak{a} be an ideal of a Noetherian ring R of characteristic $p > 0$ such that $\mathfrak{a} \cap R^\circ \neq \emptyset$, and let $N \subseteq M$ be R -modules. Given a rational number $t \geq 0$, the $t \cdot \mathfrak{a}$ -tight closure $N_M^{*t \cdot \mathfrak{a}}$ (or, \mathfrak{a}^t -tight closure $N_M^{*\mathfrak{a}^t}$, see Remark 6.2 (1) below) of N in M is defined to be the submodule of M consisting of all elements $z \in M$ for which there exists $c \in R^\circ$ such that

$$cz^q \mathfrak{a}^{[tq]} \subseteq N_M^{[q]}$$

for all $q = p^e \gg 0$, where $\lceil tq \rceil$ denotes the least integer that is greater than or equal to tq .

Remark 6.2. (1) Definition 6.1 does not change if we replace " $cz^q \mathfrak{a}^{\lceil tq \rceil} \subseteq N_M^{[q]}$ " (rounding up tq) by " $cz^q \mathfrak{a}^{\lfloor tq \rfloor} \subseteq N_M^{[q]}$ " (rounding down tq), as long as $\mathfrak{a} \cap R^\circ \neq \emptyset$. This is because the difference of $\lceil tq \rceil$ and $\lfloor tq \rfloor$ as the exponents of \mathfrak{a} is "absorbed" by the term $c \in R^\circ$. Similarly, it is easy to see that $t \cdot \mathfrak{a}^n$ -tight closure is the same as $tn \cdot \mathfrak{a}$ -tight closure for every nonnegative integer n ; cf. the proof of [HW, Proposition 2.6]. This being so, it is preferable to say " \mathfrak{a}^t -tight closure" rather than " $t \cdot \mathfrak{a}$ -tight closure." In the sequel, we always use "exponential notation" in this manner and denote the \mathfrak{a}^t -tight closure of N in M by $N_M^{*\mathfrak{a}^t}$.

(2) The above formulation extends to several rational coefficients (or, several rational exponents). Namely, given ideals $\mathfrak{a}_1, \dots, \mathfrak{a}_r \subseteq R$ with $\mathfrak{a}_i \cap R^\circ \neq \emptyset$ and rational numbers $t_1, \dots, t_r \geq 0$, if $t_i = tn_i$ for nonnegative $t \in \mathbb{Q}$ and $n_i \in \mathbb{Z}$ with $i = 1, \dots, r$, we can define $\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_r^{t_r}$ -tight closure to be $(\mathfrak{a}_1^{n_1} \cdots \mathfrak{a}_r^{n_r})^t$ -tight closure. If N is a submodule of an R -module M , then an element $z \in M$ is in the $\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_r^{t_r}$ -tight closure $N_M^{*\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_r^{t_r}}$ of N in M if and only if there exists $c \in R^\circ$ such that $cz^q \mathfrak{a}_1^{\lceil t_1 q \rceil} \cdots \mathfrak{a}_r^{\lceil t_r q \rceil} \subseteq N_M^{[q]}$ for all $q = p^e \gg 0$.

(3) We also have an analogous notion of Δ -tight closure for a pair (R, Δ) of a normal ring R and a \mathbb{Q} -Weil divisor Δ on $Y = \operatorname{Spec} R$; see [T], [HW]. If $\mathfrak{a}_i = x_i R$ and $\Delta = \sum_{i=1}^n t_i \cdot \operatorname{div}_Y(x_i)$ for $x_i \in R$ and $0 \leq t_i \in \mathbb{Q}$ with $1 \leq i \leq n$, then $\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_n^{t_n}$ -tight closure is the same as Δ -tight closure.

Definition 6.3. Let \mathfrak{a} be an ideal of a Noetherian ring R of characteristic $p > 0$ such that $\mathfrak{a} \cap R^\circ \neq \emptyset$, and let $t \geq 0$ be a rational number. We say that an element $c \in R^\circ$ is an \mathfrak{a}^t -test element if $cz^q \mathfrak{a}^{\lceil tq \rceil} \subseteq I^{[q]}$ for all $q = p^e$ whenever $z \in I^{*\mathfrak{a}^t}$.

Some results for \mathfrak{a} -tight closure generalize to those for \mathfrak{a}^t -tight closure without essential change of proofs. However, we must be careful about the difference of round-up and round-down when speaking of \mathfrak{a}^t -test elements. As a matter of fact, the following theorem is proved similarly to Theorem 1.7 (1), but the proof does not work if we replace $\lceil tq \rceil$ by $\lfloor tq \rfloor$ in Definition 6.3.

Theorem 6.4. Let R be an F -finite reduced ring of characteristic $p > 0$ and let $c \in R^\circ$ be an element such that the localized ring R_c is strongly F -regular. Then some power c^n of c is an \mathfrak{a}^t -test element for all ideals $\mathfrak{a} \subseteq R$ with $\mathfrak{a} \cap R^\circ \neq \emptyset$ and all rational numbers $t \geq 0$.

We can define the ideal $\tau(\mathfrak{a}^t)$ in a similar way to Proposition-Definition 1.9. Also, Theorem 1.13 generalizes to the case of $\tau(\mathfrak{a}^t)$ with the same proof. We summarize the results for excellent reduced local rings in the following.

Definition-Theorem 6.5. Let (R, \mathfrak{m}) be an excellent reduced local ring of characteristic $p > 0$ with $E = E_R(R/\mathfrak{m})$, and let $\mathfrak{a} \subseteq R$ be an ideal such that $\mathfrak{a} \cap R^\circ \neq \emptyset$. Given a rational number $t \geq 0$, we define the ideal $\tau(\mathfrak{a}^t) \subseteq R$ by

$$\tau(\mathfrak{a}^t) = \bigcap_M \operatorname{Ann}_R(0_M^{*\mathfrak{a}^t}) = \bigcap_{M \subseteq E} \operatorname{Ann}_R(0_M^{*\mathfrak{a}^t}) = \bigcap_{I \subseteq R} (I : I^{*\mathfrak{a}^t}),$$

where M runs through all finitely generated R -modules (resp. finitely generated R -submodules of E) in the second term (resp. the third term), and I runs through all

ideals of R . Moreover, if R is normal and \mathbb{Q} -Gorenstein, then

$$\tau(\mathfrak{a}^t) = \text{Ann}_R(0_E^{*\mathfrak{a}^t}).$$

Remark 6.6. We can define the ideal $\tau(\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_n^{t_n})$ with several rational coefficients by replacing \mathfrak{a}^t -tight closure in 6.5 by $\mathfrak{a}_1^{t_1} \cdots \mathfrak{a}_n^{t_n}$ -tight closure as defined in Remark 6.2 (2). See Theorem 6.10 (2).

Proposition 1.15 also generalizes without essential change of the proof, but we cannot replace the round-up $\lceil tq \rceil$ by the round-down $\lfloor tq \rfloor$ in the following.

Proposition 6.7. Let (R, \mathfrak{m}) be a d -dimensional excellent normal local ring of characteristic $p > 0$, $\mathfrak{a} \subseteq R$ an ideal such that $\mathfrak{a} \cap R^\circ \neq \emptyset$, and let $t \geq 0$ be a rational number. Then $0_{H_{\mathfrak{m}}^d(R)}^{*\mathfrak{a}^t}$ is the unique maximal proper submodule N with respect to the property

$$\mathfrak{a}^{\lceil tq \rceil} F^e(N) \subseteq N \text{ for all } q = p^e,$$

where $F^e: H_{\mathfrak{m}}^d(R) \rightarrow H_{\mathfrak{m}}^d(R)$ is the e -times iterated Frobenius induced on $H_{\mathfrak{m}}^d(R)$.

Now we generalize Theorem 3.4, which is the main theorem of Section 3, to the case of rational coefficients.

Theorem 6.8. Let $t \geq 0$ be a fixed rational number, R a normal \mathbb{Q} -Gorenstein local ring essentially of finite type over a field, and let \mathfrak{a} be a nonzero ideal. Assume that $\mathfrak{a} \subseteq R$ is reduced from characteristic zero to characteristic $p \gg 0$, together with a log resolution $f: X \rightarrow Y = \text{Spec } R$ of the ideal \mathfrak{a} such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$. Then

$$\tau(\mathfrak{a}^t) = H^0(X, \mathcal{O}_X(\lceil K_{X/Y} - tZ \rceil)).$$

Sketch of the proof. This is also proved in a similar way to Theorem 3.4. So we just indicate the points where some modification is needed in the following. First, we note that Lemma 3.6 holds for rational coefficients without changing the proof, i.e., $\tau((\mathfrak{a}S)^t) \cap R = \tau(\mathfrak{a}^t)$ under the assumption of Lemma 3.6, and that Lemma 3.7 is already proved for rational coefficients. Hence we can use a canonical covering of R to reduce the proof of Theorem 6.8 to the quasi-Gorenstein case. Then it is sufficient to prove the following generalization of Theorem 3.9.

Theorem 6.9. Let $t \geq 0$ be a fixed rational number, (R, \mathfrak{m}) a d -dimensional normal local ring essentially of finite type over a field, and let \mathfrak{a} be a nonzero ideal. Assume that $\mathfrak{a} \subseteq R$ is reduced from characteristic zero to characteristic $p \gg 0$, together with a log resolution $f: X \rightarrow Y = \text{Spec } R$ of \mathfrak{a} such that $\mathfrak{a}\mathcal{O}_X = \mathcal{O}_X(-Z)$. Then

$$0_{H_{\mathfrak{m}}^d(R)}^{*\mathfrak{a}^t} = \text{Ker} \left(H_{\mathfrak{m}}^d(R) \xrightarrow{\delta} H_E^d(\mathcal{O}_X(tZ)) \right),$$

where E is the closed fiber of f and δ is an edge map as in Proposition 3.8.

Here we note that the canonical dual of the sheaf $\mathcal{O}_X(tZ) = \mathcal{O}_X(\lfloor tZ \rfloor)$ is $\omega_X(-\lfloor tZ \rfloor) = \omega_X(\lceil -tZ \rceil)$, which is isomorphic to $\mathcal{O}_X(\lceil K_{X/R} - tZ \rceil)$ via $\omega_R \cong R$ if R is quasi-Gorenstein.

The inclusion $0_{H_{\mathfrak{m}}^d(R)}^{*\mathfrak{a}^t} \supseteq \text{Ker } \delta$ of the above theorem holds true in arbitrary fixed characteristic $p > 0$: Just take an element c in the proof of Proposition 3.8 from $\mathfrak{a}^{\lceil tq \rceil} \subseteq H^0(X, \mathcal{O}_X(-tqZ))$ instead of $\mathfrak{a}^q = H^0(X, \mathcal{O}_X(-qZ))$, which gives rise to a map $cF^e: H_E^d(\mathcal{O}_X(tZ)) \rightarrow H_E^d(\mathcal{O}_X(tZ))$. Then one sees that $\mathfrak{a}^{\lceil tq \rceil} F^e(\text{Ker } \delta) \subseteq \text{Ker } \delta$ for all $q = p^e$, and Proposition 6.7 applies.

To prove the reverse inclusion $0_{H_m^d(R)}^{*\mathfrak{a}^t} \subseteq \text{Ker } \delta$, we choose, in characteristic zero before reducing to characteristic $p \gg 0$, a nonzero element $c \in \mathfrak{a}$ such that R_c is regular and a log resolution $f: X \rightarrow \text{Spec } R$ of the ideal $c\mathfrak{a}$, as in the proof of Theorem 3.9. Then choose an f -ample f -exceptional \mathbb{Q} -Cartier divisor D and a sufficiently small $\varepsilon > 0$ so that $\lfloor \tilde{Z} + \varepsilon \text{div}_X(c) \rfloor = \lfloor tZ \rfloor$, where $\tilde{Z} = tZ - D$. We then move to reduction modulo $p \gg 0$ and let $m = r + 2s + \lceil n_0 t \rceil$, keeping the integers r, s, n_0 just the same as in the proof of Theorem 3.9, i.e., the ideal \mathfrak{a} is generated by r elements, c^s is a usual test element and also an \mathfrak{a}^t -test element, and $\mathcal{K} = \bigoplus_{n \geq 0} H^0(X, \omega_X(-\lfloor n\tilde{Z} \rfloor))$ is generated in degree $\leq n_0$ as a graded module over $\mathcal{R} = \bigoplus_{n \geq 0} H^0(X, \mathcal{O}_X(-n\tilde{Z}))$. Now, arguing as before, we obtain the required inclusion $0_{H_m^d(R)}^{*\mathfrak{a}^t} \subseteq \text{Ker } \delta$. \square

Finally, we note that the results in Section 4 also generalize to rational coefficients, with the same proof; see [DEL], [How], [La] for the corresponding results for multiplier ideals.

Theorem 6.10. *Let t, t' be any nonnegative rational numbers.*

- (1) (Restriction theorem): *Under the assumption of Theorem 4.1 we have*

$$\tau((\mathfrak{a}S)^t) \subseteq \tau(\mathfrak{a}^t)S.$$

- (2) (Subadditivity in regular local rings; cf. Remark 6.6): *Under the assumption of Theorem 4.5 we have*

$$\tau(\mathfrak{a}^t \mathfrak{b}^{t'}) \subseteq \tau(\mathfrak{a}^t) \tau(\mathfrak{b}^{t'}).$$

- (3) *Under the assumption of Theorem 4.8, let $\mathfrak{a} \subseteq R$ be a toric ideal. Then $\tau(\mathfrak{a}^t) = \mathcal{J}(\mathfrak{a}^t)$, and it is also a toric ideal generated by monomials x^m with $m \in M$ such that*

$$m + w \in \text{Int}(t \cdot P(\mathfrak{a})).$$

REFERENCES

- [AKM] I. Aberbach, M. Katzman, and B. MacCrimmon, *Weak F -regularity deforms in \mathbb{Q} -Gorenstein rings*, J. Algebra **204** (1998), 281–285. MR **99d**:13003
- [AM] I. Aberbach and B. MacCrimmon, *Some results on test ideals*, Proc. Edinburgh Math. Soc. (2) **42** (1999), 541–549. MR **2000i**:13005
- [B] J.-F. Boutot, *Singularités rationnelles et quotients par les groupes réductifs*, Invent. Math. **88** (1987), 65–68. MR **88a**:14005
- [BS] H. Skoda and J. Briançon, *Sur la clôture intégrale d'un idéal de germes de fonctions holomorphes en un point de \mathbb{C}^n* , C. R. Acad. Sci. Paris Sér. A **278** (1974), 949–951. MR **49**:5394
- [BH] W. Bruns and J. Herzog, *Cohen–Macaulay Rings*, Cambridge Studies in Advanced Mathematics, vol. 39, Cambridge University Press, Cambridge, 1993. MR **95h**:13020
- [DEL] J.-P. Demailly, L. Ein and R. Lazarsfeld, *A subadditivity property of multiplier ideals*, Michigan Math. J. **48** (2000), 137–156. MR **2002a**:14016
- [Ei] L. Ein, *Multiplier ideals, vanishing theorems and applications*, in Algebraic Geometry—Santa Cruz 1995, pp. 203–219, Proc. Sympos. Pure Math., vol. 62, American Mathematical Society, Providence, RI, 1997. MR **98m**:14006
- [ELS] L. Ein, R. Lazarsfeld and K. E. Smith, *Uniform bounds and symbolic powers on smooth varieties*, Invent. Math. **144** (2001), 241–252. MR **2002b**:13001
- [FW] R. Fedder and K.-i. Watanabe, *A characterization of F -regularity in terms of F -purity*, in Commutative Algebra, Berkeley 1987, pp. 227–245, Math. Sci. Res. Inst. Publ., vol. 15, Springer-Verlag, New York, 1989. MR **91k**:13009

- [GS] S. Goto and Y. Shimoda, *On the Rees algebras of Cohen-Macaulay local rings*, in *Commutative Algebra, Fairfax 1979*, pp. 201–231, Lecture Notes in Pure and Appl. Math., vol. 68, Dekker, New York, 1982. MR **84a**:13021
- [GW] S. Goto and K.-i. Watanabe, *On graded rings, I*, J. Math. Soc. Japan **30** (1978), 179–213. MR **81m**:13021
- [Ha1] N. Hara, *A characterization of rational singularities in terms of injectivity of Frobenius maps*, Amer. J. Math. **120** (1998), 981–996. MR **99h**:13005
- [Ha2] ———, *Geometric interpretation of tight closure and test ideals*, Trans. Amer. Math. Soc. **353** (2001), 1885–1906. MR **2001m**:13009
- [HT] N. Hara and S. Takagi, *Some remarks on a generalization of test ideals*, preprint.
- [HW] N. Hara and K.-i. Watanabe, *F-regular and F-pure rings vs. log terminal and log canonical singularities*, J. Algebraic Geometry **11** (2002), 363–392. MR **2002k**:13009
- [HWY1] N. Hara, K.-i. Watanabe, and K. Yoshida, *F-rationality of Rees algebras*, J. Algebra **247** (2002), 153–190.
- [HWY2] ———, *Rees algebras of F-regular type*, J. Algebra **247** (2002), 191–218.
- [HHK] M. Herrmann, E. Hyry and T. Korb, *On Rees algebras with a Gorenstein Veronese subring*, J. Algebra **200** (1998), 279–311. MR **98m**:13006
- [Ho1] M. Hochster, *Cyclic purity versus purity in excellent Noetherian rings*, Trans. Amer. Math. Soc. **231** (1977), 463–488. MR **57**:3111
- [Ho2] M. Hochster, *The tight integral closure of a set of ideals*, J. Algebra **230** (2000), 184–203. MR **2002f**:13009
- [HH0] M. Hochster and C. Huneke, *Tight closure and strong F-regularity*, Mem. Soc. Math. France **38** (1989), 119–133. MR **91i**:13025
- [HH1] ———, *Tight closure, invariant theory, and the Briançon-Skoda theorem*, J. Amer. Math. Soc. **3** (1990), 31–116. MR **91g**:13010
- [HH2] ———, *F-regularity, test elements, and smooth base change*, Trans. Amer. Math. Soc. **346** (1994), 1–62. MR **95d**:13007
- [HH3] ———, *Tight closure in equal characteristic zero*, to appear.
- [How] J. A. Howald, *Multiplier ideals of monomial ideals*, Trans. Amer. Math. Soc. **353** (2001), 2665–2671. MR **2002b**:14061
- [Hu] C. Huneke, *Tight closure and its applications*, C.B.M.S. Regional Conference Series in Mathematics, No. 88, American Mathematical Society, Providence, RI, 1996. MR **96m**:13001
- [Hy1] E. Hyry, *Blow-up rings and rational singularities*, Manuscripta Math. **98** (1999), 377–390. MR **2001d**:13002
- [Hy2] ———, *Coefficient ideals and the Cohen-Macaulay property of Rees algebras*, Proc. Amer. Math. Soc. **129** (2001), 1299–1308. MR **2001h**:13005
- [Ka] Y. Kawamata, *The cone of curves of algebraic varieties*, Ann. Math. **119** (1984), 603–633. MR **86c**:14013b
- [Ku] E. Kunz, *On Noetherian rings of characteristic p* , Amer. J. Math. **98** (1976), 999–1013. MR **55**:5612
- [La] R. Lazarsfeld, *Positivity in Algebraic Geometry*, preprint.
- [Li] Lipman, J., *Adjoints of ideals in regular local rings*, Math. Res. Letters **1** (1994), 739–755. MR **95k**:13028
- [Mc] B. MacCrimmon, *Weak F-regularity is strong F-regularity for rings with isolated non-Q-Gorenstein points*, Trans. Amer. Math. Soc., to appear.
- [Ma] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics, vol. 8, Cambridge University Press, Cambridge, 1986. MR **88h**:13001
- [MS] V. B. Mehta and V. Srinivas, *A characterization of rational singularities*, Asian J. Math. **1** (1997), 249–278. MR **99e**:13009
- [N] A. Nadel, *Multiplier ideal sheaves and Kähler-Einstein metrics of positive scalar curvature*, Ann. Math. **132** (1990), 549–596. MR **92d**:32038
- [Si] A. K. Singh, *F-regularity does not deform*, Amer. J. Math. **121** (1999), 919–929. MR **2000e**:13006
- [Sm1] K. E. Smith, *F-rational rings have rational singularities*, Amer. J. Math. **119** (1997), 159–180. MR **97k**:13004
- [Sm2] ———, *The multiplier ideal is a universal test ideal*, Comm. Algebra **28** (2000), 5915–5929. MR **2002d**:13008

- [T] S. Takagi, *An interpretation of multiplier ideals via tight closure*, preprint.
- [VV] P. Valabrega and G. Valla, *Form rings and regular sequences*, Nagoya Math. J. **72** (1978), 93–101. MR **80d**:14010
- [Vr] A. Vraciu, *Strong test ideals*, J. Pure Appl. Algebra **167** (2002), 361–373. MR **2003a**:13004
- [Wa] K.-i. Watanabe, *F-regular and F-pure normal graded rings*, J. Pure Appl. Algebra **71** (1991), 341–350. MR **92g**:13003
- [Wi] L. J. Williams, *Uniform stability of kernels of Koszul cohomology indexed by the Frobenius endomorphism*, J. Algebra **172** (1995), 721–743. MR **96f**:13003

MATHEMATICAL INSTITUTE, TOHOKU UNIVERSITY, SENDAI 980-8578, JAPAN

E-mail address: hara@math.tohoku.ac.jp

GRADUATE SCHOOL OF MATHEMATICS, NAGOYA UNIVERSITY, CHIKUSA-KU, NAGOYA 464-8602, JAPAN

E-mail address: yoshida@math.nagoya-u.ac.jp

SESHADRI CONSTANTS ON JACOBIAN OF CURVES

JIAN KONG

ABSTRACT. We compute the Seshadri constants on the Jacobian of hyperelliptic curves, as well as of curves with genus three and four. For higher genus curves we conclude that if the Seshadri constants of their Jacobian are less than 2, then the curves must be hyperelliptic.

1. INTRODUCTION AND STATEMENT OF RESULTS

Let X be a smooth complex projective variety. Let L be an ample line bundle. Let $p \in X$ be a point. Define the Seshadri constant of L at p to be the real number

$$\epsilon(L, p) := \inf \left\{ \frac{C \cdot L}{\text{mult}_p C} \mid p \in C \subset X \right\}.$$

Here the infimum is taken over all reduced curves C passing through p , and $\text{mult}_p C$ is the multiplicity of C at p . Another equivalent definition is

$$\epsilon(L, p) = \sup \{ \epsilon \mid f^*L - \epsilon E \text{ is nef} \},$$

where $f : Bl_p X \rightarrow X$ is the blow-up of X at p and E is the exceptional divisor.

The Seshadri constant indicates how far the ample divisor is from the boundary of the ample cone near point p , and thus measures positivity, or ampleness locally. The study of Seshadri constants has drawn increasing interest during recent years. For properties of Seshadri constants see [1] and [6].

In the case of abelian varieties, it is known that a general element in the moduli space of principally polarized abelian varieties of dimension g has Seshadri constant very close to its maximum upper bound ([4]). On the other hand, there are some special abelian varieties, namely Jacobian, which have relatively small Seshadri constants. We will discuss some cases in this paper.

Let C be a smooth projective algebraic curve over \mathbf{C} with genus $g = g(C) \geq 2$. Denote Θ to be the theta divisor of $J(C)$, its Jacobian (recall $J(C) = \text{Pic}^0(C)$). Since abelian varieties are homogeneous spaces, we can define $\epsilon = \epsilon(\Theta, 0) = \epsilon(\Theta, p)$ for any $p \in J(C)$.

It is known that $1 < \epsilon \leq \sqrt{g}$, and if C is hyperelliptic, then $\epsilon \leq \frac{2g}{g+1}$ ([4], [5]). In particular, if $g = 2$, then C is hyperelliptic and it is known that $\epsilon = \frac{4}{3}$ ([6]). The problem becomes very interesting even when $g = 3$. The point here is to see if the Seshadri constants can be their maximum, i.e., \sqrt{g} , thus most of the time irrational, or always less than their maximum – and thus more likely rational. While all the

Received by the editors August 1, 2002 and, in revised form, August 26, 2002.

2000 *Mathematics Subject Classification*. Primary 14H40; Secondary 14K12.

Key words and phrases. Algebraic geometry, algebraic curves, abelian varieties.

existing examples suggest the latter, we investigate this problem in detail, mainly by looking at the cases when $\epsilon \leq 2$.

Our main result is the following theorem:

Theorem 1.1. *Assume the Picard number of $J(C)$ is one. Then:*

- (1) *If C is hyperelliptic, then $\epsilon = \frac{2g}{g+1}$.*
- (2) *If $g = 3$ and C is not hyperelliptic, then $\epsilon = \frac{12}{7}$.*
- (3) *If $g = 4$ and C is not hyperelliptic, then $\epsilon = 2$.*
- (4) *If $g \geq 5$ and C is not hyperelliptic, then $\epsilon \geq 2$.*

Part (4) of the theorem can be restated as:

Corollary 1.2. *If $g \geq 5$ and $\epsilon < 2$, then C is hyperelliptic.*

Remark 1.3. (1) For the ease of calculation on the Neron-Severi group of the symmetric product C_2 , we need that it is generated by a fiber and the diagonal, i.e., its Picard number is 2. That is true if C is of general moduli. We need this condition throughout this paper. But this restriction, however, seems to be not essential.

(2) We can also locate all the special curves that give relatively small ratios in cases (1) to (3).

2. PROOF OF THEOREM: HYPERELLIPTIC CASE

The following observation, while straightforward, points out where we want to find special curves that give the exact value of the Seshadri constants.

Lemma 2.1. *If C' is an irreducible curve in $J(C)$ such that $\frac{C' \cdot \Theta}{\text{mult}_0 C'} \leq 2$, then for any divisor D with $D \equiv k\Theta$ and $\text{mult}_0 D \geq 2k$, we have $C' \subset D$.*

If C is hyperelliptic, then the case of $k = 1$ in Lemma 2.1 reads $D \equiv \Theta$ and $\text{mult}_0 D \geq 2$, which we denote as (*). For $d \geq 2$, let C_d be the d -fold symmetric product of C (the set of effective divisors of degree d on C).

Proposition 2.2. *Let $u : C_d \rightarrow J(C)$ be the Abel-Jacobi map. Then*

$$\bigcap_{(*)} D = u(C_2).$$

Proof. Let L be a hyperelliptic line bundle on C . Let p_0 be a ramification point of the g_2^1 ; so $L = \mathcal{O}_C(2p_0)$. We fix a translation of the Abel-Jacobi map $u : C_d \rightarrow J(C)$ by sending $Y \in C_d$ to $Y - \deg(Y) \cdot p_0 \in J(C)$, and for simplicity we ignore the p_0 part for representation of points in $J(C)$ in our proof. Recall that $\phi : C_{g-3} \rightarrow C_{g-1}$, $Y \mapsto Y + L$ maps C_{g-3} birationally and surjectively to $\text{Sing}(\Theta)$ in this case ([3]).

For any $Y \in C_{g-3}$, define $D_Y = \Theta - Y$. It translates $Y + 2p_0 \in \text{Sing}(\Theta)$ to $0 \in D_Y$. Thus $D_Y \equiv \Theta$ and $\text{mult}_0 D_Y \geq 2$. It suffices to show that $\bigcap D_Y = u(C_2)$.

It is obvious that $u(C_2) \subset \bigcap D_Y$, since for any point $(p, q) \in C_2$ we can rewrite it as $(p + q + Y) - Y \in D_Y$ for any $Y \in C_{g-3}$.

On the other side, any points in $\bigcap D_Y$ can be represented as $D - Y$ for some $D \in C_{g-1}$ and $Y \in C_{g-3}$. Also, since it is in the intersection, for any $F \in C_{g-3}$, there exists $E \in C_{g-1}$ such that $D - Y = E - F$, i.e., $D - Y + F$ is (equivalent to) an effective divisor for any F . We claim $D - Y$ itself must be effective, and since it has degree 2, it is in $u(C_2)$.

Pick a representation of $D - Y$ such that Y contains no ramification point of g_2^1 . First assume that D contains no hyperelliptic pair. If $D - Y$ is not effective, pick $p \in Y$ but $p \notin D$, and let $L = \mathcal{O}_C(p + p')$. Choose $F = Y - p + p'$. Then $D - Y + F = D - p + p'$. But the linear system $|D - p + p'|$ is empty, since otherwise it must contain a multiple of hyperelliptic pairs and base points, which will lead to $p \in D$.

If D has some hyperelliptic pairs, then cancel as many points in $D - Y$ as possible until either $D - Y$ is effective or D runs out of hyperelliptic pairs and reduce to a similar situation in the first case. \square

If C is hyperelliptic, and $\text{rk}(NS(C_2)) = 2$, then $NS(C_2)$ is generated by F and Δ , the image of a fiber and the diagonal from the natural map $C \times C \rightarrow C_2$. There is a rational curve, call it \mathbf{P}^1 , that consists of hyperelliptic pairs $\{(p, q) \in C_2 \mid \mathcal{O}_C(p + q) = L\}$. Also denote $u^*(\Theta)$ still as Θ . We list the numerical properties of $NS(C_2)$ below.

Lemma 2.3. *With notation as above, we have:*

- (1) $\Theta = (g + 1)F - \frac{1}{2}\Delta$ and $\mathbf{P}^1 = 2F - \frac{1}{2}\Delta$.
- (2) $F^2 = 1, F \cdot \Delta = 2, \Delta^2 = 4 - 4g$.

The Abel-Jacobi map $u : C_2 \rightarrow J(C)$ contracts \mathbf{P}^1 and is isomorphic outside of \mathbf{P}^1 . Now let C'' be an irreducible curve in C_2 not contracted by u and let $C' = u(C'')$. Then

$$\frac{C'' \cdot \Theta}{C'' \cdot \mathbf{P}^1} = \frac{C' \cdot \Theta}{\text{mult}_0 C'}.$$

So our theorem in the hyperelliptic case follows from the following proposition.

Proposition 2.4. *Among all irreducible curves in C_2 not contracted by u , Δ is the only curve with minimum ratio $\frac{\Delta \cdot \Theta}{\Delta \cdot \mathbf{P}^1} = \frac{2g}{g+1}$.*

Proof. Since $\Delta \cdot \Theta = 4g$ and $\Delta \cdot \mathbf{P}^1 = 2g + 2$, we have $\frac{\Delta \cdot \Theta}{\Delta \cdot \mathbf{P}^1} = \frac{4g}{2g+2} = \frac{2g}{g+1}$.

Let $C_0 = aF + b\Delta \subset C_2$ be an irreducible curve not contracted by u . Then $C_0 \cdot \mathbf{P}^1 = a + (2g + 2)b \geq 0$ and $C_0 \cdot \Theta = (a + 4b)g > 0$. Then

$$\frac{\Delta \cdot \Theta}{\Delta \cdot \mathbf{P}^1} = \frac{(a + 4b)g}{a + (2g + 2)b} > \frac{2g}{g + 1} \iff a > 0.$$

But if $a \leq 0$, then we must have $b > 0$ since $C_0 \cdot \Theta = (a + 4b)g > 0$. Now we have $C_0 \cdot \Delta = 2a + b(4 - 4g) < 0$. Since both C_0 and Δ are irreducible, $C_0 = \Delta$. \square

Remark 2.5. A little more detailed calculation shows that Δ is actually the only curve whose corresponding ratio is less than two.

3. PROOF OF THE THEOREM: NON-HYPERELLIPTIC CASE

If C is non-hyperelliptic, then choose the case $k = 2$ in Lemma 2.1 which reads $D \equiv 2\Theta$ and $\text{mult}_0 D \geq 4$. Denote this linear system as $|2\Theta|_{00}$. So we look at the base locus of $|2\Theta|_{00}$. Here we need the following result of Welters.

Proposition 3.1 (Welters [7]). *$Bs(|2\Theta|_{00}) = \lambda(C \times C)$. Here $\lambda : C \times C \rightarrow J(C)$, $\lambda(p, q) = p - q$, is the difference map.*

Remark 3.2. Welters' theorem is true for all curves with $g = 3$ or $g \geq 5$. For $g = 4$ the base locus has two more isolated points, which will not affect our proof since we are looking at curves inside the base locus.

In this case we look at the Neron-Severi group in $C \times C$. It is generated by fibers F_1, F_2 and the diagonal Δ . We list their numerical properties below.

Lemma 3.3. *With notation as above, we have:*

- (1) $\lambda^*\Theta = (g-1)(F_1 + F_2) + \Delta$.
- (2) $F_i^2 = 0, F_1 \cdot F_2 = F_i \cdot \Delta = 1, \Delta^2 = 2 - 2g, i = 1, 2$.

Since C is non-hyperelliptic, the difference map λ contracts the diagonal Δ to $0 \in J(C)$ and is isomorphic outside Δ . This enables us, as similarly in the hyperelliptic case, to shift the computation from the ratio $\frac{C \cdot \Theta}{\text{mult}_0 C}$ in $J(C)$ to the ratio of the intersection numbers in the Neron-Severi group, which are well understood. Specifically, let C'' be an irreducible curve in $C \times C$ not contracted by λ and let $C' = u(C'')$. Then

$$\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta} = \frac{C' \cdot \Theta}{\text{mult}_0 C'}.$$

So our theorem in the non-hyperelliptic case follows from the following proposition.

Proposition 3.4. *With notation as above:*

- (1) If $g = 3$, the minimum ratio $\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta}$ is $\frac{12}{7}$ for curves in $C \times C$, and is achieved by one curve.
- (2) If $g = 4$, the minimum ratio $\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta}$ is 2 for curves in $C \times C$, and is achieved by more than one curve.
- (3) If $g \geq 5$, then $\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta} \geq 2$ for all curves in $C \times C$ not contracted by λ .

Proof. (1) $g = 3$: In this case, the canonical system embeds C as a plane quartic. Let $\mathcal{O}_C(1)$ be its hyperplane section. Consider the curve

$$C_0 = \{(p, q) \mid \mathcal{O}_C(p + q + 2r) = \mathcal{O}_C(1) \text{ for some } r \in C\} \subset C_2.$$

Write $C_0 = aF + b\Delta$. Since $C_0 \cdot \Delta = 56$ (twice the number of bitangents) and $C_0 \cdot F = 10$ (degree of the ramification divisor of the dual curve's g_3^1), we can solve for a and b and get $C_0 = 16F - 3\Delta$. C_0 is irreducible since it is isomorphic to C via $p + q \rightarrow r$. Pulling it back to $C \times C$ we get a curve $C_0'' = 16(F_1 + F_2) + 6\Delta$. Now

$$\frac{C_0'' \cdot \lambda^*\Theta}{C_0'' \cdot \Delta} = \frac{[16(F_1 + F_2) - 6\Delta] \cdot [2(F_1 + F_2) + \Delta]}{[16(F_1 + F_2) - 6\Delta] \cdot \Delta} = \frac{96}{56} = \frac{12}{7}.$$

To claim that $\frac{12}{7}$ is the minimum ratio, let $C'' = aF_1 + bF_2 + c\Delta$ be any irreducible curve in $C \times C$ not contracted by λ . If $C'' \neq C_0''$, then $C'' \cdot C_0'' = 10(a+b) + 56c \geq 0$. So if $c \geq 0$, then

$$\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta} = \frac{3(a+b)}{a+b-4c} \geq 3 > \frac{12}{7}.$$

If $c < 0$, then

$$\frac{C'' \cdot \lambda^*\Theta}{C'' \cdot \Delta} = \frac{3(a+b)}{a+b-4c} \geq \frac{3(a+b)}{a+b+\frac{5}{7}(a+b)} = \frac{7}{4} > \frac{12}{7}.$$

This shows that the only curve that achieves the minimum ratio $\frac{12}{7}$ is C_0'' .

(2) $g = 4$: In this case C has two g_3^1 's. Let L be one g_3^1 . Consider the curve $C_0 = \{(p, q) \mid |L - p - q| > 0\} \subset C_2$. Since $C_0 \cdot F = 2$ and $C_0 \cdot \Delta = 12$ (degree

of ramification divisor of L), we find that $C_0 = 3F - \frac{1}{2}\Delta$. Lift to $C \times C$ to get $C_0'' = 3(F_1 + F_2) - \Delta$. A similar calculation as above shows that

$$\frac{C_0'' \cdot \lambda^* \Theta}{C_0'' \cdot \Delta} = \frac{24}{12} = 2,$$

and it is the minimum ratio that can be achieved on $C \times C$.

Note that in this case there is another curve (from the other g_3^1) that gives the minimum ratio. The reason is that in this case $C_0^2 = 0$, while in the case of $g = 3$ we have $C_0^2 < 0$ (thus unique).

(3) $g \geq 5$: Assume C has a g_d^1 ($d \geq 3$), call it L . As in (2), consider the curve $C_0 = \{(p, q) \mid |L - p - q| > 0\} \subset C_2$. Then $C_0 \cdot F = d - 1$ and $C_0 \cdot \Delta = 2d + 2g - 2$ (degree of ramification divisor of L). Thus $C_0 = dF - \frac{1}{2}\Delta$. Lifting to $C \times C$ we get $C_0'' = d(F_1 + F_2) - \Delta$. Now first we have

$$\frac{C_0'' \cdot \lambda^* \Theta}{C_0'' \cdot \Delta} = \frac{dg}{d + g - 1} > 2.$$

Secondly, for any irreducible $C'' = aF_1 + bF_2 + c\Delta \subset C \times C$ not contracted by λ , either $C'' \cdot C_0'' < 0$, or

$$\frac{C'' \cdot \lambda^* \Theta}{C'' \cdot \Delta} \geq \frac{d + g - 1}{d} \geq 2 \text{ if } d \leq g - 1.$$

Since the Brill-Noether number for g_d^1 is non-negative if $d \geq \frac{g+2}{2}$, both ratios above are at least 2. If the minimum ratio $\frac{C'' \cdot \lambda^* \Theta}{C'' \cdot \Delta} < 2$, then all the curves C_0 must be reducible and contain an irreducible component C_1 whose lift to $C \times C$ gives a small ratio. It is easy to see that $C_1^2 < 0$, thus unique in C_2 . This is certainly impossible. (For example, if there are two different g_d^1 for some d , then to have a common component for corresponding $C_0 \subset C_2$, one coordinate has to be a base point of g_d^1 . Thus it is a linear combination of fibers, and since the component is irreducible, it is a fiber. But for a fiber, the corresponding ratio is $g > 2$.)

Note that the minimum ratio exists and can be achieved if $\frac{dg}{d+g-1} \leq \frac{d+g-1}{d}$, which is equivalent to $\frac{dg}{d+g-1} \leq \sqrt{g}$, or $d \leq \sqrt{g} + 1$. \square

4. OTHER RELATED PROBLEMS OF SESHADRI CONSTANTS

For non-hyperelliptic cases when $g \geq 5$, to find the Seshadri constants, the first step is to look at the curves in C_2 . It is related to the problem whether the cone of effective curves of C_2 is closed. If it is, the curve from the boundary will give a better upper bound of $\epsilon(\Theta)$ which is less than \sqrt{g} . In all special cases that we have discussed (hyperelliptic, small genus, curves with g_d^1 for small d), the cone is closed. For the general case there is some indication that the curve from one boundary (the other one being the diagonal), if closed, will give the ratio $\frac{gg}{p}$ where (p, q) is the primitive solution of Pell's equation $x^2 - gy^2 = 1$, if g is not a square. The following example gives some indication that it could be true.

Example 4.1. If C is a plane quintic (i.e., genus is 6), consider the curve $C_0 = \{(p, q) \mid |\mathcal{O}_C(1) - p - q - 2r| > 0\} \subset C_2$. Then C_0 is irreducible and $C_0 = 50F - 7\Delta$. Since any curve $C' \subset C_2$ satisfies $C_0 \cdot C' \geq 0$, a calculation shows that on $C \times C$, $\frac{C'' \cdot \lambda^* \Theta}{C'' \cdot \Delta} \geq \frac{12}{5}$ holds for all irreducible curves (except Δ). Note that the bound $\frac{12}{5}$ is

what the conjecture gives. Also note that one expects small values for plane curves that are special in the moduli of curves, so general curves of genus 6 must also satisfy that bound.

For $g \geq 5$, if $\epsilon < 2$, then it follows that C is hyperelliptic. In all cases, hyperelliptic curves give us the smallest Seshadri constants. From the known result ([2]) of $B_s(|2\Theta|_{00})$ in dimension 4, it is easy to see that:

If A is an indecomposable principally polarized abelian variety of dimension 4 and $\epsilon(\Theta) < 2$, then A is the Jacobian of a hyperelliptic curve C of genus 4.

It is very reasonable to ask the same question for any genus and seems like it could be true.

ACKNOWLEDGEMENT

The author thanks Aaron Bertram for his encouragement and help.

REFERENCES

1. L. Ein and R. Lazarsfeld, Seshadri constants on smooth surfaces, *Journées de Géométrie Algébrique d'Orsay (Orsay, 1992)*, Astérisque 218 (1993), 177-186. MR **95f**:14031
2. E. Izadi, The geometric structure of A_4 , the structure of the Prym map, double solids and Γ_{00} -divisors, *J. Reine Angew. Mathematik* 462 (1995), 93-158. MR **96d**:14042
3. H. Lange and C. Birkenhake, *Complex abelian varieties*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 1992. MR **94j**:14001
4. R. Lazarsfeld, Lengths of periods and Seshadri constants of abelian varieties, *Math. Res. Letters* 3 (1996), 439-447. MR **98e**:14044
5. M. Nakamaye, Seshadri constants on abelian varieties, *Amer. J. Math.* 118 (1996), 621-635. MR **97k**:14005
6. A. Steffens, Remarks on Seshadri constants, *Math. Z.* 227 (1998), 505-510. MR **99c**:14009
7. G. Welters, The surfaces $C - C$ on Jacobi varieties and second order theta functions, *Acta Math.* 157 (1986), 1-22. MR **87j**:14048

DEPARTMENT OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY, BALTIMORE, MARYLAND 21218
E-mail address: jkong@math.jhu.edu

ON THE CLIFFORD ALGEBRA OF A BINARY FORM

RAJESH S. KULKARNI

ABSTRACT. The Clifford algebra C_f of a binary form f of degree d is the k -algebra $k\{x, y\}/I$, where I is the ideal generated by $\{(\alpha x + \beta y)^d - f(\alpha, \beta) \mid \alpha, \beta \in k\}$. C_f has a natural homomorphic image A_f that is a rank d^2 Azumaya algebra over its center. We prove that the center is isomorphic to the coordinate ring of the complement of an explicit Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$, where C is the curve $(w^d - f(u, v))$ and g is the genus of C .

1. INTRODUCTION

Let f be a form of degree d in n variables over a field k . Then the Clifford algebra C_f is the k -algebra $k\{x_1, \dots, x_n\}/I$ where $k\{x_1, \dots, x_n\}$ is the free associative algebra in n variables and I is the ideal generated by $\{(\alpha_1 x_1 + \dots + \alpha_n x_n)^d - f(\alpha_1, \dots, \alpha_n) \mid \alpha_1, \dots, \alpha_n \in k\}$. If $d = 2$, this is the classical Clifford algebra of a quadratic form. If $d > 2$, then this is sometimes called the *generalized* Clifford algebra and has been studied by various authors, including Roby [21], Revoy [20], and Childs [3].

The first case of higher degree (that is, $d = 3, n = 2$) was studied by Haile in [8] and [9]. He showed that the Clifford algebra of a binary cubic form over a field k , with characteristic $\neq 2, 3$, is Azumaya with center the affine coordinate ring of an elliptic curve. Further, this elliptic curve is the Jacobian of the projective curve given by the equation $(w^3 - f(u, v))$.

The case of $d > 3$ displays different behavior. Namely, using results of [20], it is easy to see that C_f contains a free algebra on two variables. In particular, C_f is not finitely generated over its center (as a module) and hence is not Azumaya. However, it has been shown ([10]) that the dimension of any representation is divisible by d . Furthermore, $\tilde{C}_f = C_f / (\bigcap \ker \eta)$ is Azumaya over its center, where the intersection is taken over all the kernels of dimension d representations. The question of describing the center is thus interesting from this point of view.

The main tool in understanding representations of the Clifford algebra (for $d \geq 3$) was introduced by Van den Bergh in [22]. While previous results about representations ([8], [10]) were global in nature, this idea gave a local result. To understand the statement we introduce some notation. Denote by C the projective curve given by the equation $(w^d - f(u, v))$. Let g be the genus of this degree d curve. Assume that k is algebraically closed of characteristic 0 and that the form f has no repeated factors over k . By an rd -dimensional representation of C_f , we mean a homomorphism from C_f to $M_{rd}(k)$, the $rd \times rd$ matrices over k . Then the result

Received by the editors January 1, 2002.

2000 *Mathematics Subject Classification*. Primary 16H05, 16G99, 14H40, 14K30.

([22], Proposition 1 in Section 1 and Lemma 2 in Section 2.2) states that there is a one-to-one correspondence between equivalence classes of rd -dimensional representations of C_f and isomorphism classes of vector bundles \mathcal{E} on C of rank r and degree $(d + g - 1)$ such that $H^0(C, \mathcal{E}(-1)) = 0$. It should be emphasized that this result is constructive in nature. Namely, for a vector bundle that satisfies the conditions of the theorem, we can construct an explicit representation of C_f (on the global sections of \mathcal{E}), and vice versa.

This result suggests that the center of \tilde{C}_f should be the coordinate ring of an affine open set in the (translated) Jacobian of the curve C . In fact, a closer inspection reveals that the open set is the complement of a Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$. The results of [8] and [9] pertaining to the case of binary cubic forms give credence to this belief. In this paper, we prove this result under the assumptions that k is an infinite field, f has no repeated factors over an algebraic closure of k and $\text{char}(k)$ does not divide d .

First consider the case when C has a k -rational point. For this case, we briefly describe the strategy to construct the required isomorphism. Let Z_k denote the center of \tilde{C}_f and A_k denote the coordinate ring of the complement of a Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$. We compute the graded module associated to the direct image of the universal bundle under the projection $C \times_k \text{Spec } A_k \rightarrow \mathbb{P}_k^1 \times_k \text{Spec } A_k$. In fact, it turns out that this direct image is isomorphic to the pull-back of the direct image of the universal bundle under the projection on the second factor. This allows us to define a homomorphism φ from \tilde{C}_f to $\text{End}_{A_k} P_0$, where P_0 is the A_k -module of global sections of the direct image of the universal line bundle. This morphism has the important property that any d -dimensional representation of \tilde{C}_f factors through $\text{End}_{A_k} P_0$. This implies that the morphism φ is injective. Using the fact that A_k is integrally closed, we prove that φ maps Z_k to A_k . Then our goal is to show that this morphism on the centers is an isomorphism.

In our original approach ([12]), we showed that φ is an isomorphism by proving sufficiently many geometric properties of φ to use the Main Theorem of Zariski. However, it was pointed out to us that a more direct approach might be to construct a morphism η from $\text{Spec } Z_k$ to $\text{Spec } A_k$ using some of the ideas from [23]. Namely, we construct a natural transformation between functors which are represented by the schemes $\text{Spec } Z_k$ and $\text{Spec } A_k$. This then yields the required morphism. It is then easy to show that the composition $\eta \circ \varphi : \text{Spec } A_k \rightarrow \text{Spec } A_k$ is the identity morphism. Using the injectivity of $\varphi : Z_k \rightarrow A_k$, we see that this is an isomorphism.

Now we discuss the case when C does not have a k -rational point. In this case, there exists a finite Galois extension k'/k such that $C(k')$ is nonempty. The results discussed in the previous paragraphs show that after this base extension, the center $Z_{k'}$ is isomorphic (via φ) to $A_{k'}$, the coordinate ring of the complement of the Θ -divisor in $\text{Pic}_{C'/k'}^{d+g-1}$. We show that this morphism descends to k . For any element $\sigma \in \text{Gal}(k'/k)$, consider the automorphism $\mu = \sigma^{-1} \circ \varphi^{-1} \circ \sigma \circ \varphi$. Using a calculation of the universal bundle \mathcal{P} , we show that the restriction of its pull-back under μ to a closed point y is isomorphic to \mathcal{P}_y . This then shows that μ is the identity morphism, giving us the required descent.

The paper is organized in six sections. In the second section, we review some basic material about Jacobians of curves, keeping in mind the audience for this paper. In Section 3, we study the restriction of the direct image of the universal

line bundle on the complement of the Θ -divisor. In Section 4, the morphism φ is constructed and its properties are studied. In Section 5, we construct the morphism η . In the last section, the main theorem is proved, identifying the center of \tilde{C}_f .

In the following, we assume that the binary form f has no repeated factors over an algebraic closure of k and that the characteristic of k does not divide d .

NOTATION AND CONVENTIONS

- All rings have an identity element.
- Let B be an A -algebra and X be a $\text{Spec } A$ -scheme. We denote $X \times_{\text{Spec } A} \text{Spec } B$ by $X \times_A B$.
- All schemes are locally Noetherian and all morphisms are of locally finite type.
- (Sch/k) denotes the category of schemes over $\text{Spec } k$ whose structure morphisms are locally of finite type. Also $(\text{Sch}/X)_{\text{fl}}$ denotes the flat site on X . All sheafifications are with respect to the flat topology.
- (fppf) = faithfully-flat and of finite presentation.
- The projections of fibred products onto the i^{th} component X_i are denoted by either p_i, π_i or p_{X_i}, π_{X_i} .
- For any scheme X and a positive integer n , $X^{(n)}$ denotes the n -fold symmetric product of X .
- The terms line bundles (respectively vector bundles) and invertible sheaves (respectively locally free sheaves) are used interchangeably.
- For any vector bundle \mathcal{E} , $\chi(\mathcal{E})$ denotes the Euler characteristic of \mathcal{E} .
- The translation morphism induced by an element a of an abelian variety A is denoted by t_a .
- All the fields are assumed to be infinite.

2. THE UNIVERSAL PROPERTY OF THE PICARD SCHEME AND THE UNIVERSAL INVERTIBLE SHEAF

In this section, we recall some basic properties of Picard schemes of curves and the universal invertible sheaf associated with them. The general references for this section are [2], Chapter 8, and [15].

Let S be a base scheme, X an S -scheme and $f : X \rightarrow S$ be the structure morphism. For any S -scheme T , let q denote the projection $X \times_S T \rightarrow T$. The functor $\text{Pic}_{X/S}$ which associates to any S -scheme T the Picard group $\text{Pic}(X \times_S T)$ is called the Picard functor. This functor is not representable, since it is not a sheaf even with respect to the Zariski topology. We consider its sheafification with respect to the (fppf) -topology (called the *relative Picard functor*). The sheafified functor is representable by a scheme which is also denoted by $\text{Pic}_{X/S}$. If f has a section, the sheafified functor is isomorphic to the functor that assigns to any S -scheme T the group $\text{Pic}(X \times_S T)/q^*\text{Pic}(T)$ (Proposition 4 in [2], Chapter 8).

There is another description for the relative Picard functor for more restricted situations ([2], Chapter 8). We assume that $f_*(\mathcal{O}_X) = \mathcal{O}_S$ holds universally (that is, this formula holds true after any base change) and that f admits a section $\varepsilon : S \rightarrow X$. For a line bundle \mathcal{L} on X , an isomorphism $\alpha : \mathcal{O}_S \xrightarrow{\sim} \varepsilon^*(\mathcal{L})$ is called a *rigidification* of \mathcal{L} . The pair (\mathcal{L}, α) is referred to as a rigidified line bundle along the section ε . There is an obvious notion of morphisms of rigidified line bundles. Then we consider the functor $(P, \varepsilon) : (\text{Sch}/S)^0 \rightarrow (\text{Sets})$ which associates to an

S -scheme T the set of isomorphism classes of line bundles on $X_T = X \times_S T$ that are rigidified along the induced section $\varepsilon_T : T \rightarrow X_T$. The functor (P, ε) is canonically isomorphic to the relative Picard functor $\text{Pic}_{X/S}$.

Now consider a smooth projective curve C over a field k of genus g . We call the open (and closed) subfunctor that considers only invertible sheaves \mathcal{L} on $C \times_k T$ whose restriction to $C \times \{t\}$ is of fixed degree n for any point t the *relative Picard functor of degree n* . The scheme that represents it is denoted by $\text{Pic}_{C/k}^n$. This is a smooth, projective scheme over k .

Next suppose the curve C has a k -rational point. Fix a section $\varepsilon : \text{Spec } k \rightarrow C$ corresponding to such a rational point. Let $\mathcal{O}(1)$ be a very ample line bundle on C of degree d . Then the scheme $\text{Pic}_{C/k}$ (respectively $\text{Pic}_{C/k}^{d+g-1}$) also represents (P, ε) (respectively (P^{d+g-1}, ε) , which is an open subfunctor of (P, ε) consisting of rigidified line bundles of degree $(d+g-1)$). So the identity on $\text{Pic}_{C/k}$ (respectively $\text{Pic}_{C/k}^{d+g-1}$) gives a line bundle \mathcal{P} (respectively \mathcal{P}) on $C \times_k \text{Pic}_{C/k}$ (respectively $C \times_k \text{Pic}_{C/k}^{d+g-1}$) that is canonically rigidified along the induced section. The sheaf \mathcal{P} is called the *universal* (or *Poincaré*) *line bundle* for $(C/k, \varepsilon)$. In fact, it is easy to see that the corresponding universal line bundle on $C \times_k \text{Pic}_{C/k}^{d+g-1}$ is the pull-back of the universal line bundle on $C \times_k \text{Pic}_{C/k}$ under the canonical inclusion. The next proposition, which is adapted from [2], Chapter 8, 8.4, Proposition 4, justifies this terminology.

Proposition 2.1. *With the notation as above, the functor (P, ε) is representable by a scheme which we denote by $\text{Pic}_{C/k}$. The universal line bundle \mathcal{P} has the following property: For any k -scheme X , and for any line bundle \mathcal{L} on $C \times_k X$ that is rigidified along the induced section ε_X , there exists a unique morphism $g : X \rightarrow \text{Pic}_{C/k}$ such that \mathcal{L} , as a rigidified line bundle, is isomorphic to the pull-back of \mathcal{P} under the morphism $\text{id}_C \times g$. A similar statement is true for (P^{d+g-1}, ε) if the line bundle $\mathcal{L}|_{C \times_k \{x\}}$ is of degree $(d+g-1)$ for any point $x \in X$.*

Next we want to relate the schemes representing the relative Picard functors corresponding to C/k and C_K/K . Here K is an arbitrary field extension of k and $C_K = C \times_k K$. The next proposition is true for arbitrary base extensions as well. See [7], Section 3.

Let $\pi : C \rightarrow \text{Spec } k$ be the structure morphism. Let K be any field extension of k and $f : \text{Spec } K \rightarrow \text{Spec } k$ be the corresponding base extension. We continue to use the same notation as above in the following proposition.

Proposition 2.2. *The Picard scheme $\text{Pic}_{C_K/K}$ is isomorphic to $\text{Pic}_{C/k} \times_k K$ as a K -scheme. Let $g : \text{Pic}_{C_K/K} \rightarrow \text{Pic}_{C/k} \times_k K$ be such an isomorphism, and let \mathcal{P} be the universal line bundle for $(C/k, \varepsilon)$. Then $h_1^* h_2^* (\text{id} \times g)^* (p_1 \times \text{id})^* \mathcal{P}$ is the universal line bundle for $(C_K/K, f \circ \varepsilon)$, where the morphisms are described in the following sequence:*

$$C_K \times_K \text{Pic}_{C_K/K} \xrightarrow{h_1} C \times_k K \times_K \text{Pic}_{C_K/K} \xrightarrow{h_2} C \times_k K \times_k \text{Pic}_{C/k}.$$

Remark 2.3. The first part of the above proposition remains true even if the curve C/k does not have a k -rational point. Namely, the formation of the Picard scheme is compatible with the base change. This follows because, in fact, the usual construction of the Picard scheme is to construct it after a base extension so that the curve has a rational point (over the extended field), and then to use descent. For

example, see [15]. Also, the above proposition remains true if we restrict to the open subfunctor of degree $(d+g-1)$. This is clear since the degree of a line bundle does not change under a pull-back via base extension of fields.

Remark 2.4. We have considered in the above discussion only the open subfunctors corresponding to degree $(d+g-1)$. But all the facts about representability remain true for any degree. The facts about the universal line bundle also remain true as long as the curve under consideration has a rational point over its base field. The proofs of these assertions are identical to the case of degree $(d+g-1)$.

The Picard scheme is a group scheme (variety) over the base field. The Picard scheme of degree 0 is the Jacobian variety of C/k and is also denoted by J . The group operation corresponds to tensoring of line bundles. We define specific Θ -divisors in Picard schemes of degrees $(g-1)$ and $(d+g-1)$.

Definition 2.5. Let C/k be a curve with the hypothesis as above (we do not need C to have a k -rational point). The Θ -divisor (of degree $(g-1)$) is the schematic image of the canonical morphism

$$(C)^{(g-1)} \rightarrow \mathrm{Pic}_{C/k}^{g-1}, \quad D_T \mapsto [D_T],$$

where, for any k -scheme T and for any T -valued point D_T of $(C)^{(g-1)}$, $[D_T]$ denotes the element of $\mathrm{Pic}_{C/k}^{g-1}$ corresponding to D_T .

Recall that the curve C is equipped with a (very ample) line bundle $\mathcal{O}(1)$ of degree d . This gives a canonically defined Θ -divisor in $\mathrm{Pic}_{C/k}^{d+g-1}$, as the image of the Θ -divisor under the morphism

$$\mathrm{Pic}_{C/k}^{g-1} \xrightarrow{\otimes \mathcal{O}(1)} \mathrm{Pic}_{C/k}^{d+g-1}.$$

Now let K be a field extension of k . By considering the commutative diagram

$$\begin{array}{ccc} (C_K)^{(g-1)} & \longrightarrow & \mathrm{Pic}_{C_K/K}^{g-1} \\ \downarrow & & \downarrow \mu \\ (C)^{(g-1)} & \longrightarrow & \mathrm{Pic}_{C/k}^{g-1} \end{array}$$

we see that the pull-back of the Θ -divisor under the morphism μ is the Θ -divisor for C_K/K . In fact, the same is true as well for the complement of the Θ -divisor.

Proposition 2.6. Let C/k be a curve as before, and let K be a field extension of k . Let \mathcal{L} be a rigidified line bundle on C_K of degree $(d+g-1)$. Then the image of the unique morphism $\mathrm{Spec} K \xrightarrow{i} \mathrm{Pic}_{C/k}^{d+g-1}$ corresponding to \mathcal{L} lies in the complement of the Θ -divisor if and only if $h^0(C_K, \mathcal{L}(-1)) = 0$.

Proof. One direction is immediate once we note that the fibre over a closed point x under the composite morphism

$$C^{(g-1)} \rightarrow \mathrm{Pic}_{C/k}^{g-1} \rightarrow \mathrm{Pic}_{C/k}^{d+g-1}$$

either is empty or is exactly the linear system $H^0(C \times_k k(x), \mathcal{L}_x(-1))$, where \mathcal{L}_x is the line bundle corresponding to the point x in $\mathrm{Pic}_{C/k}^{d+g-1}$. This idea can be formalized by using the Poincaré bundle.

For the other direction, if $h^0(C_K, \mathcal{L}(-1)) \neq 0$, then we may choose a global section s of $\mathcal{L}(-1)$. Then the divisor D_s of this section gives an element of $C^{(g-1)}$. Using the canonical morphism, this gives an invertible sheaf \mathcal{L}_2 on C_K that has a global section whose divisor is D_s . Furthermore, the image of $\text{Spec } K$ corresponding to \mathcal{L}_2 is by definition in the Θ -divisor. But since their global sections have the same divisors, the invertible sheaves $\mathcal{L}(-1)$ and \mathcal{L}_2 must be isomorphic. This gives the other direction. \square

We record an easy corollary for future reference.

Corollary 2.7. *With the hypothesis as above, the pull-back (or the inverse image) of the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$ under the canonical morphism*

$$\text{Pic}_{C_K/K}^{d+g-1} \longrightarrow \text{Pic}_{C/k}^{d+g-1}$$

is the complement of the Θ -divisor in $\text{Pic}_{C_K/K}^{d+g-1}$.

Remark 2.8. Note that our idea of the proof of the corollary depends on C having a k -rational point. However, by Remark 2.3, the same question can be asked if C does not have a k -rational point. We consider a finite Galois extension k'/k (with Galois group G) so that $C(k')$ is nonempty. Let $C' = C \times_k k'$. Then $\text{Pic}_{C'/k'}^{d+g-1}$ has a G -action and gives, by descent, $\text{Pic}_{C/k}^{d+g-1}$. The morphism

$$(C')^{(g-1)} \xrightarrow{h_{k'}} \text{Pic}_{C'/k'}^{d+g-1}$$

is G -equivariant, and so the image of $h_{k'}$ is G -invariant, which descends to the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$. Hence the complement of the Θ -divisor in $\text{Pic}_{C'/k'}^{d+g-1}$ descends to the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$.

2.1. The universal line bundle on $C \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}$. In this section, we would like to explicitly construct the universal line bundle on $C \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}$. Though we consider only curves of degree d in \mathbb{P}^2 , most of the calculations remain valid for arbitrary curves. Our goal is to find a formula which relates the universal line bundle to its pull-back under a Galois automorphism of the base field.

We start by recalling the construction of the universal line bundle for degree 0. Let C be a smooth curve over a field k such that $C' = C \times_k k'$ has a k' -rational point. We assume that k'/k is a finite Galois extension. In our case, this follows from the assumption that the characteristic of k does not divide d . To see this, note that the curve given by $(w^d - f(u, v))$ has a rational point over the splitting field of the polynomial $x^d - f(a, b)$ for any elements $a, b \in k$. Let J denote the Jacobian variety $\text{Pic}_{C'/k'}^0$. Fix a rational point P on C' . Denote by Θ_0 the theta divisor in J obtained by translation of the usual Θ -divisor in $\text{Pic}_{C'/k'}^{g-1}$ by $\mathcal{L}(-(g-1)P)$. This is the same as the divisor of $\mathcal{L}(\Theta)(-1) \otimes \mathcal{L}(-(g-1)P)$, where Θ is the theta divisor in $\text{Pic}_{C'/k'}^{d+g-1}$ defined earlier. We have the morphism ([15], Section 2)

$$h : C' \longrightarrow J,$$

which maps a closed point Q to $\mathcal{L}([Q - P])$. For any divisor D in the Jacobian variety, we denote by $\mathcal{L}'(D)$ the line bundle $m^*\mathcal{L}(D) \otimes p^*\mathcal{L}(D)^{-1} \otimes q^*\mathcal{L}(D)^{-1}$ on $J \times_{k'} J$, where m is the multiplication morphism on J and p, q are the projections onto the first and second factors, respectively. Finally, write Θ_0^- for the image of

Θ_0 under the morphism $(-1)_J : J \rightarrow J$, and $(\Theta_0)_a$ for $t_a \Theta_0 = \Theta_0 + a$. Denote $(\Theta_0^-)_a$ by $(\Theta_0)_a^-$.

With this notation, Lemma 6.8 of [15] says the following:

Lemma 2.9. *The sheaf $\mathcal{L} = (h \times (-1)_J)^* \mathcal{L}'(\Theta_0^-)$ is isomorphic to the universal invertible sheaf given by the universal property of J .*

Using the line bundle of the last lemma, we construct the universal line bundle for $C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}$. This construction is well known; see, for example, [1], Chapter 4, Lemma 2.2 and the discussion preceding it. Denote by L_0 the line bundle

$$\mathcal{L}(-(g-1)P)(-1)$$

on C' , and by p_1 the projection of $C' \times_{k'} J$ onto the first factor. The morphism η is defined as

$$\begin{array}{ccc} \text{Pic}_{C'/k'}^{d+g-1} & \xrightarrow{\eta} & J \\ L & \mapsto & L \otimes L_0. \end{array}$$

We can now state the lemma describing the universal bundle of degree $(d+g-1)$.

Lemma 2.10. *With the above notation, the line bundle $\mathcal{P} = (\text{id}_{C'} \times \eta)^* \mathcal{L} \otimes p_1^*(L_0^{-1})$ is the universal line bundle of degree $(d+g-1)$.*

Proof. Let ε denote the section of $C' \rightarrow k'$ and let $\varepsilon_{\text{Pic}_{C'/k'}^{d+g-1}}, \varepsilon_J$ denote the induced sections. Then it follows that

$$(\text{id}_{C'} \times \eta) \circ \varepsilon_{\text{Pic}_{C'/k'}^{d+g-1}} = \varepsilon_J \circ \eta.$$

This means that $(\text{id}_{C'} \times \eta)^* \mathcal{L}$ is rigidifiable. Also, $p_1^*(L_0^{-1})$ is rigidifiable along $\varepsilon_{\text{Pic}_{C'/k'}^{d+g-1}}$. Thus \mathcal{P} is a rigidifiable line bundle. By Remark 2.15, this is sufficient, and we may make any choice of rigidification.

Now let T be any k' -scheme and let \mathcal{L}_T be a rigidified line bundle on $C' \times_{k'} T$ such that for any $t \in T$, the restriction $(\mathcal{L}_T)_t$ is of degree $(d+g-1)$. Then $\mathcal{L}'_T = \mathcal{L}_T \otimes p_1^*(L_0^{-1})$ is a rigidified line bundle on $C' \times_{k'} T$ such that its restriction to any C'_t is of degree 0. This determines a unique map

$$\mu : T \longrightarrow J,$$

and hence a map $\eta^{-1} \circ \mu : T \rightarrow \text{Pic}_{C'/k'}^{d+g-1}$. Using the universal property of J and \mathcal{L} and the fact that the projection p_1 commutes with $(\text{id}_{C'} \times \eta)$, we see that the line bundle $(\text{id}_{C'} \times (\eta^{-1} \circ \mu))^* \mathcal{P}$ is isomorphic to \mathcal{L}_T . Furthermore, the morphism $\eta^{-1} \circ \mu$ is unique with respect to this property. This is the universal property of \mathcal{P} . \square

Now let σ be in $\text{Gal}(k'/k)$. If this induces an automorphism of a k' -scheme, then it is also denoted by σ . For future purposes, we want to compute $\sigma^* \mathcal{P} \otimes \mathcal{P}^{-1}$. First we prove a computational lemma.

Lemma 2.11. *With the notation as before, the following commutation relations hold:*

- (1) $\sigma \circ t_{(g-2)[P^{\sigma-1}-P]} \circ m \circ (h \times (-1)_J) \circ (\text{id} \times \eta) = m \circ (h \times (-1)_J) \circ (\text{id} \times \eta) \circ \sigma,$
- (2) $\sigma \circ t_{(-1)[P^{\sigma-1}-P]} \circ p \circ (h \times (-1)_J) \circ (\text{id} \times \eta) = p \circ (h \times (-1)_J) \circ (\text{id} \times \eta) \circ \sigma,$
- (3) $\sigma \circ t_{(g-1)[P^{\sigma-1}-P]} \circ q \circ (h \times (-1)_J) \circ (\text{id} \times \eta) = q \circ (h \times (-1)_J) \circ (\text{id} \times \eta) \circ \sigma.$

Proof. We prove only the first relation, since the others are proved in a similar way. Let μ_l (respectively μ_r) denote the left (respectively the right) side of the first equation. Let $\mu = m \circ (\mu_l, (-1) \circ \mu_r)$. Fix a closed point Q on C' and a divisor D associated to a line bundle that corresponds to a closed point in $\text{Pic}_{C'/k'}^{d+g-1}$. Let E denote the divisor of a section of $\mathcal{L}((g-1)P)(1)$. Then

$$\begin{aligned} \mu(Q, D) &= \mu_l(Q, D) - \mu_r(Q, D) \\ &= \sigma \circ t_{(g-2)[P^{\sigma-1}-P]} \circ m \circ (h \times (-1)_J)(Q, D - E) \\ &\quad - m \circ (h \times (-1)_J) \circ (\text{id} \times \eta)(Q^\sigma, D^\sigma) \\ &= \sigma \circ t_{(g-2)[P^{\sigma-1}-P]} \circ m([Q - P], E - D) \\ &\quad - m \circ (h \times (-1)_J)(Q^\sigma, D^\sigma - E) \\ &= \sigma \circ t_{(g-2)[P^{\sigma-1}-P]}([Q - P] + E - D) - m \circ (Q^\sigma - P, E - D^\sigma) \\ &= [Q^\sigma - P^\sigma + E^\sigma - D^\sigma + (g-2)(P - P^\sigma)] - [Q^\sigma - P + E - D^\sigma], \end{aligned}$$

which is the empty divisor. (This follows from the observation that we may take $E^\sigma - E = (g-1)[P^\sigma - P]$.) This means that $\mu(Q, D)$ is the trivial line bundle. So, by the rigidity lemma ([14], Theorem 2.1), the two morphisms are the same. \square

Now let \mathcal{P} be the universal bundle on $C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}$, and let σ be as above.

Lemma 2.12. *Let x be a closed point in $\text{Pic}_{C'/k'}^{d+g-1}$ and $i : \text{Spec } k(x) \rightarrow \text{Pic}_{C'/k'}^{d+g-1}$ be the corresponding morphism. Then, with the above notation, we have that the line bundle $(\text{id}_{C'} \times i)^*(\sigma^*\mathcal{P} \otimes \mathcal{P}^{-1})$ on $C' \times k(x)$ is trivial. In other words, $(\sigma^*\mathcal{P} \otimes \mathcal{P}^{-1})$ is isomorphic to p_2^*L , where L is a line bundle on $\text{Pic}_{C'/k'}^{d+g-1}$ and p_2 is the projection onto the second factor.*

Proof. We first compute $\sigma^*(\text{id}_{C'} \times \eta)^*\mathcal{L} \otimes (\text{id}_{C'} \times \eta)^*\mathcal{L}^{-1}$. To this end, consider $\sigma^*(\text{id}_{C'} \times \eta)^*\mathcal{L}$. Using the definition of \mathcal{L} and the relations of the last lemma, we have that

$$\begin{aligned} \sigma^*(\text{id}_{C'} \times \eta)^*\mathcal{L} &= \sigma^*(\text{id}_{C'} \times \eta)^*(h \times (-1)_J)^*\mathcal{L}'(\Theta_0^-) \\ &= \sigma^*(\text{id}_{C'} \times \eta)^*(h \times (-1)_J)^*(m^*\mathcal{L}(\Theta_0^-) \otimes p^*\mathcal{L}(\Theta_0^-)^{-1} \otimes q^*\mathcal{L}(\Theta_0^-)^{-1}) \\ &= (\text{id}_{C'} \times \eta)^*(h \times (-1)_J)^*(m^*\mathcal{L}((\Theta_0^-)_{-(g-2)[P^{\sigma-1}-P]+(g-1)[P^{\sigma-1}-P]}^-) \\ &\quad \otimes p^*\mathcal{L}(-(\Theta_0^-)_{[P^{\sigma-1}-P]+(g-1)[P^{\sigma-1}-P]}^-) \\ &\quad \otimes q^*\mathcal{L}(-(\Theta_0^-)_{-(g-1)[P^{\sigma-1}-P]+(g-1)[P^{\sigma-1}-P]}^-)) \\ &= (\text{id}_{C'} \times \eta)^*(h \times (-1)_J)^*(m^*\mathcal{L}((\Theta_0^-)_{[P^{\sigma-1}-P]}^-) \\ &\quad \otimes p^*\mathcal{L}(-(\Theta_0^-)_{g[P^{\sigma-1}-P]}^-) \otimes q^*\mathcal{L}(-(\Theta_0^-)^-)). \end{aligned}$$

Now consider the line bundle $m^*\mathcal{L}((\Theta_0^-)_{[P^{\sigma-1}-P]}^- - \Theta_0^-)$. By [14], Proposition 10.1, the line bundle $\mathcal{L}((\Theta_0^-)_{[P^{\sigma-1}-P]}^- - \Theta_0^-)$ lies in $\text{Pic}^0(\text{Pic}_{C'/k'}^0)$. But, by [14], Proposition 9.2, for any line bundle \mathcal{L}' on $\text{Pic}_{C'/k'}^0$ that lies in $\text{Pic}^0(\text{Pic}_{C'/k'}^0)$, we have

$m^*\mathcal{L}' \cong p^*\mathcal{L}' \otimes q^*\mathcal{L}'$. Combining this fact with the above calculation, we have

$$\begin{aligned} & \sigma^*(\mathrm{id}_{C'} \times \eta)^*\mathcal{L} \otimes (\mathrm{id}_{C'} \times \eta)^*\mathcal{L}^{-1} \\ &= (\mathrm{id}_{C'} \times \eta)^*(h \times (-1)_J)^* \\ & \quad (p^*\mathcal{L}((\Theta_0)_{[P^{\sigma^{-1}}-P]}^- - (\Theta_0)_{g[P^{\sigma^{-1}}-P]}^-) \otimes q^*\mathcal{L}((\Theta_0)_{[P^{\sigma^{-1}}-P]}^- - \Theta_0^-)). \end{aligned}$$

Recall that we are interested in the restriction of this line bundle to $C' \times_{k'} k(x)$. Since

$$q \circ (h \times (-1)_J) \circ (\mathrm{id}_{C'} \times \eta) \circ (\mathrm{id}_{C'} \times i) = (-1)_J \circ \eta \circ i \circ q,$$

it follows that the pull-back of the q^* -component of the above line bundle to $C' \times_{k'} k(x)$ is trivial, and so we may consider only the p^* -component. Note that

$$p \circ (h \times (-1)_J) \circ (\mathrm{id}_{C'} \times \eta) \circ (\mathrm{id}_{C'} \times i) = h \circ p_1.$$

This gives us

$$\begin{aligned} & (\mathrm{id}_{C'} \times i)^*(\sigma^*(\mathrm{id}_{C'} \times \eta)^*\mathcal{L} \otimes (\mathrm{id}_{C'} \times \eta)^*\mathcal{L}^{-1}) \\ (2.1) \quad & \cong p_1^*h^*\mathcal{L}((\Theta_0)_{[P^{\sigma^{-1}}-P]}^- - (\Theta_0)_{g[P^{\sigma^{-1}}-P]}^-). \end{aligned}$$

Now let $h^{(g)}$ denote the morphism $(C')^{(g)} \rightarrow \mathrm{Pic}_{C'/k'}^0$, which maps an effective divisor D of degree g to the line bundle associated to $[D - gP]$. Then it is clear that $h^{(g)}(P^{\sigma^{-1}} + (g-1)P) = [P^{\sigma^{-1}} - P]$ and $h^{(g)}(gP^{\sigma^{-1}}) = g[P^{\sigma^{-1}} - P]$. So by [15], Lemma 6.8, the invertible sheaf $h^*\mathcal{L}((\Theta_0)_{[P^{\sigma^{-1}}-P]}^-)$ is isomorphic to $\mathcal{L}([P^{\sigma^{-1}} + (g-1)P])$, and $h^*\mathcal{L}((\Theta_0)_{g[P^{\sigma^{-1}}-P]}^-) \cong \mathcal{L}(g[P^{\sigma^{-1}}])$. Using (2.1) and these remarks, we get

$$\begin{aligned} & (\mathrm{id}_{C'} \times i)^*(\sigma^*(\mathrm{id}_{C'} \times \eta)^*\mathcal{L} \otimes (\mathrm{id}_{C'} \times \eta)^*\mathcal{L}^{-1}) \\ (2.2) \quad & \cong p_1^*\mathcal{L}(P^{\sigma^{-1}} + (g-1)P - gP^{\sigma^{-1}}). \end{aligned}$$

Now we consider the line bundle $(\mathrm{id}_{C'} \times i)^*(\sigma^*\mathcal{P} \otimes \mathcal{P}^{-1})$ on $C' \times_{k'} k(x)$. To finish the lemma, we consider the line bundle $p_1^*(L_0)^{-1}$. (Note that this p_1 is the projection of $C' \times \mathrm{Pic}_{C'/k'}^{d+g-1}$ onto the first factor. We continue with this notation to avoid complicating our notation even further.) It follows from the definition of L_0 that

$$\begin{aligned} & (\mathrm{id}_{C'} \times i)^*(\sigma^*p_1^*(L_0)^{-1} \otimes p_1^*(L_0)) \\ (2.3) \quad & \cong (\mathrm{id}_{C'} \times i)^*\mathcal{L}((g-1)(P^{\sigma^{-1}} - P)) \\ & \cong p_1^*\mathcal{L}((g-1)(P^{\sigma^{-1}} - P)). \end{aligned}$$

Combining (2.2), (2.3) and the definition of \mathcal{P} , we see that the line bundle

$$(\mathrm{id}_{C'} \times i)^*(\sigma^*\mathcal{P} \otimes \mathcal{P}^{-1})$$

on $C' \times_{k'} k(x)$ is trivial.

The last part of the statement now follows from the Seesaw Theorem, Cor. 6, Sect. 5, Chap. 2 in [18]. \square

2.2. Auxiliary lemmas about rigidifications of line bundles. The following lemma will be used repeatedly in later sections.

Lemma 2.13. *Let C/k be a curve as before with a k -rational point, $\varepsilon : \operatorname{Spec} k \rightarrow C$. Let K be a field extension of k . Then any two line bundles $\mathcal{L}_1, \mathcal{L}_2$ on C_K that are rigidified along ε_K are isomorphic as rigidified line bundles if they are isomorphic as line bundles.*

Another lemma, which will be used later, concerns line bundles on $C \times_k D$, where $D = k[t]/(t^2)$ is the ring of dual numbers. However, for later purposes, it is sufficient to consider the case when k is algebraically closed.

Fix a section $\varepsilon : \operatorname{Spec} k \rightarrow C$. We denote the induced section $\operatorname{Pic}_{C/k}^{d+g-1} \rightarrow C \times_k \operatorname{Pic}_{C/k}^{d+g-1}$ by $\varepsilon_{\operatorname{Pic}_{C/k}^{d+g-1}}$. Let x be a closed point of $\operatorname{Pic}_{C/k}^{d+g-1}$, and let

$$t_x : \operatorname{Spec} D \longrightarrow \operatorname{Pic}_{C/k}^{d+g-1}$$

be a tangent vector at x . This gives a rigidified line bundle \mathcal{L}_1 on $C \times_k D$. Suppose that \mathcal{L}_2 is another line bundle on $C \times_k D$ such that we have an isomorphism

$$\eta : \mathcal{L}_1 \longrightarrow \mathcal{L}_2$$

of line bundles. Further suppose that \mathcal{L}_2 is isomorphic (as a line bundle) to $j^* \mathcal{L}'_2$ for some line bundle \mathcal{L}'_2 on C , where j is the morphism $\operatorname{Spec} D \rightarrow \operatorname{Spec} k$ corresponding to the natural morphism $k \hookrightarrow D$.

Lemma 2.14. *Under the above hypothesis and notation, there exists a rigidification on \mathcal{L}'_2 such that \mathcal{L}_1 and $j^* \mathcal{L}'_2$ are isomorphic as rigidified line bundles. Here $j^* \mathcal{L}'_2$ is rigidified using the rigidification on \mathcal{L}'_2 in a natural way.*

Remark 2.15. We omit proofs of these lemmas. The first lemma can be proved by comparing the set of rigidifications of \mathcal{L}_2 with the automorphism group of \mathcal{L}_2 . The second lemma can be proved by comparing the sets of rigidifications of \mathcal{P} , $j^* \mathcal{L}'_2$ and \mathcal{L}'_2 with automorphism groups of $\mathcal{O}_{\operatorname{Pic}_{C/k}^{d+g-1}}$, \mathcal{O}_D and \mathcal{O}_k . However, the proofs are easier if we use yet another definition of the relative Picard functor. From [2], Section 8.1, the relative Picard functor is isomorphic to the functor which associates to an S -scheme T the group $\operatorname{Pic}(X_T)/\operatorname{Pic}(T)$. In fact, this shows that two rigidifiable line bundles on X_T that are isomorphic are isomorphic as rigidified line bundles.

3. THE COMPLEMENT OF THE Θ -DIVISOR AND THE UNIVERSAL LINE BUNDLE

In this section, we discuss the complement of the Θ -divisor. However, after some generalities, we consider only those curves which are of interest from the point of view of the Clifford algebra. In particular, the projection of the restriction of the universal bundle to $C \times_k$ (complement of the Θ -divisor) on $\mathbb{P}_k^1 \times_k$ (complement of the Θ -divisor) will be considered in detail. (The projection $C \rightarrow \mathbb{P}_k^1$ is described below.) We continue with the notation and definitions from the last section.

We first prove that the complement of the Θ -divisor is an affine scheme. This is an immediate consequence of well-known properties of the Θ -divisor.

Proposition 3.1. *Let C/k be a degree d , genus g smooth curve in \mathbb{P}^2 . (C need not have a k -rational point.) Then the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$ is an affine scheme.*

Proof. First note that we may assume that C has a k -rational point. This is because the Θ -divisor and its complement were obtained by Galois descent on $\text{Pic}_{C'/k'}^{d+g-1}$, where k'/k is an extension so that $C(k')$ is nonempty.

Suppose C has a k -rational point. Then the Θ -divisor is ample in $\text{Pic}_{C/k}^{d+g-1}$. See [15], Remark 6.5 and Theorem 6.6. Now the proposition follows, since the complement of any ample divisor in $\text{Pic}_{C/k}^{d+g-1}$ is an affine scheme. \square

Remark 3.2. If C has a k -rational point, then the same proof shows that the complement of any Θ -divisor in $\text{Pic}_{C/k}^m$ is affine for any m .

Notation. For any curve C/k as above, we denote by $\text{Spec } A_k$ the open complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$.

For the rest of this section we assume that C has a k -rational point.

Recall from the introduction that we are mainly interested in the curve $C = \text{Proj}(k[u, v, w]/(w^d - f(u, v)))$. (Here f is the given binary form.) We prove that in the cases of interest, the curve C is always nonsingular.

Lemma 3.3. *The curve C is nonsingular provided that f does not have repeated factors over an algebraic closure of k and that the characteristic of k does not divide d .*

Proof. We may assume that k is algebraically closed. An easy computation shows that

$$\frac{\partial(w^d - f(u, v))}{\partial u} = -\frac{\partial f}{\partial u}, \quad \frac{\partial(w^d - f(u, v))}{\partial v} = -\frac{\partial f}{\partial v}$$

and that $\frac{\partial(w^d - f(u, v))}{\partial w} = dw^{d-1}$. Also we have that $u \frac{\partial g}{\partial u} + v \frac{\partial g}{\partial v} = dg$ for any polynomial $g(u, v)$. Since the characteristic of k does not divide d , C is singular only if the system of equations $\{f, \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v}\}$ has a solution. However, since f has no repeated factors, this is impossible. \square

Corresponding to the inclusion $k[u, v] \rightarrow (k[u, v, w]/(w^d - f(u, v)))$ we have the projection

$$C \xrightarrow{q_k} \mathbb{P}_k^1.$$

For any field extension K/k , we denote the morphism obtained as above by q_K . Let y be a closed point in $\text{Spec } A_k$, and let $k(y)$ be its residue field. Then we denote by \mathcal{L}_y the pull-back of the universal line bundle \mathcal{P} under the canonical morphism

$$C \times_k k(y) \longrightarrow C \times_k \text{Pic}_{C/k}^{d+g-1}.$$

The following proposition is a key proposition for our purposes and was proved in [22] when k is algebraically closed. We give a proof for the sake of completeness.

Proposition 3.4. *The coherent sheaf $\mathcal{F}_y = (q_{k(y)})_* \mathcal{L}_y$ is isomorphic to the trivial vector bundle $\bigoplus_d \mathcal{O}_{\mathbb{P}_{k(y)}^1}$.*

Proof. We first assume that k is algebraically closed. Since \mathbb{P}_k^1 is a nonsingular curve over an algebraically closed field, any torsion-free coherent sheaf is a vector bundle. Since $(q_k)_*$ respects torsion-freeness ([6] I, Proposition 8.4.1), \mathcal{F}_y is a vector bundle on \mathbb{P}_k^1 . Since any vector bundle on \mathbb{P}_k^1 is a sum of line bundles, we can write $\mathcal{F}_y \cong \bigoplus_{i=1}^d \mathcal{O}_{\mathbb{P}_k^1}(n_i)$ for some integer n_i 's. Note that $\chi(\mathcal{L}_y) = \chi(\mathcal{F}_y)$. Then, by [5], Cor. 15.2.1,

$$\begin{aligned} \chi(\mathcal{F}_y) &= (\text{rank}(\mathcal{F}_y))(1 - g_{\mathbb{P}_k^1}) + \deg(\mathcal{F}_y), \\ (1 - g_C) + \deg(\mathcal{L}_y) &= d + \sum_i n_i. \end{aligned}$$

Here $\deg(\mathcal{E})$ denotes the degree of a vector bundle \mathcal{E} . This gives $\sum_i n_i = 0$. Since y is in the complement of the Θ -divisor, we have

$$\begin{aligned} h^0(\mathbb{P}_k^1, (q_k)_*(\mathcal{L}_y \otimes_{\mathcal{O}_C} \mathcal{O}_C(-1))) &= h^0(\mathbb{P}_k^1, (q_k)_*(\mathcal{L}_y(-1))) \\ &= h^0(C, \mathcal{L}_y(-1)), \end{aligned}$$

which is zero. But by the projection formula, we have the identity

$$(q_k)_*(\mathcal{L}_y \otimes_{\mathcal{O}_C} (q_k^* \mathcal{O}_{\mathbb{P}_k^1}(-1))) \cong (q_k)_* \mathcal{L}_y \otimes_{\mathcal{O}_{\mathbb{P}_k^1}} \mathcal{O}_{\mathbb{P}_k^1}(-1).$$

Using this and the previous equations gives

$$h^0(\mathbb{P}_k^1, \bigoplus_i \mathcal{O}_{\mathbb{P}_k^1}(n_i - 1)) = 0;$$

so $n_i \leq 0$ for all $i = 1, \dots, d$, and hence $n_i = 0$ (since $\sum_i n_i = 0$).

So $(q_k)_* \mathcal{L}_y \cong \bigoplus_d \mathcal{O}_{\mathbb{P}_k^1}$. Now let k be any field, and let y be a point as before. Then we consider the (canonically rigidified) pull-back of the line bundle \mathcal{L}_y under the canonical morphism $C \times_k \bar{k} \xrightarrow{j_C} C \times_k k(y)$. Then by Corollary 2.7, the image of the point corresponding to $j_C^* \mathcal{L}_y$ lies in $\text{Spec } A_{\bar{k}}$. By the earlier part of the argument, $(q_{\bar{k}})_*(j_C^*(\mathcal{L}_y)) = \bigoplus_d \mathcal{O}_{\mathbb{P}_{\bar{k}}^1}$. To finish the proof we consider the following commutative diagram:

$$(3.1) \quad \begin{array}{ccc} C_{\bar{k}} & \xrightarrow{j_C} & C_{k(y)} \\ q_{\bar{k}} \downarrow & & \downarrow q_{k(y)} \\ \mathbb{P}_{\bar{k}}^1 & \xrightarrow{j_{\mathbb{P}^1}} & \mathbb{P}_{k(y)}^1 \end{array}$$

Now $(q_{\bar{k}})_*(j_C^*(\mathcal{L}_y)) \cong (j_{\mathbb{P}^1}^*)(q_{k(y)})_* \mathcal{L}_y$ by [11], Chapter III, Theorem 9.3. But $\text{Aut}((j_{\mathbb{P}^1}^*)(q_{k(y)})_* \mathcal{L}_y) = \text{GL}_d(\bar{k})$ and $H^1(\text{Gal}_{\bar{k}/k(y)}, \text{GL}_d(\bar{k})) = 0$ by Hilbert 90. So it follows that $\mathcal{F}_y \cong \bigoplus_d \mathcal{O}_{\mathbb{P}_{k(y)}^1}$. □

Next we want to consider the direct image of the universal line bundle on $C \times_k \text{Pic}_{C/k}^{d+g-1}$ under the projection onto the second factor. In general, there is no reason for this to be a vector bundle. For our purposes, the restriction of this direct image to $\text{Spec } A_k$ is more important. With the help of the last proposition we prove the following.

Proposition 3.5. *Let $\pi : C \times_k \text{Pic}_{C/k}^{d+g-1} \rightarrow \text{Pic}_{C/k}^{d+g-1}$ denote the projection onto the second factor. Then $(\pi_* \mathcal{P})|_{\text{Spec } A_k}$ is a locally free sheaf of rank d .*

Proof. First note that since $\text{Spec } A_k \rightarrow \text{Pic}_{C/k}^{d+g-1}$ is an open immersion,

$$\pi_*(\mathcal{P} |_{C \times_k \text{Spec } A_k}) \cong (\pi_* \mathcal{P}) |_{\text{Spec } A_k}.$$

For any point y in $\text{Spec } A_k$, let $k(y)$ be the residue field of y . We denote by \mathcal{P}_y the line bundle $(\text{id} \times i)^* \mathcal{P}$ on $C_{k(y)}$, where i is the morphism $\text{Spec } k(y) \rightarrow \text{Pic}_{C/k}^{d+g-1}$. We will prove that $\dim_{k(y)} H^0(C_{k(y)}, \mathcal{P}_y)$ is constant. Since $\text{Spec } A_k$ is an open set in an integral scheme $\text{Pic}_{C/k}^{d+g-1}$, it is itself irreducible and reduced. Then by [18], Corollary 2, p. 50, it follows that $(\pi_* \mathcal{P}) |_{\text{Spec } A_k}$ is a locally free sheaf of rank d .

First we consider a closed point y in $\text{Spec } A_k$. Then, with the notation from diagram (3.1), $(q_{k(y)})_* \mathcal{P}_y$ is a trivial rank d vector bundle. So

$$h^0(C_{k(y)}, \mathcal{P}_y) = h^0(\mathbb{P}_{k(y)}^1, (q_{k(y)})_* \mathcal{P}_y) = d.$$

This holds for all closed points in $\text{Spec } A_k$. By the upper semicontinuity of the function $y \mapsto \dim_{k(y)} H^0(C_{k(y)}, \mathcal{P}_y)$ and the fact that $\text{Spec } A_k$ is a Jacobson scheme (in particular, any open set contains a closed point), it follows that $\dim_{k(y)} H^0(C_{k(y)}, \mathcal{P}_y)$ is constant. \square

The locally free sheaf on $\text{Spec } A_k$ in the statement of the previous proposition will be denoted by \mathcal{E} . As part of the proof of the last proposition we proved that the function $y \mapsto \dim_{k(y)} H^0(C_{k(y)}, \mathcal{P}_y)$ is constant. This, combined with [18], Corollary 2, p. 50, gives the following corollary.

Corollary 3.6. *For all $y \in \text{Spec } A_k$, the natural map*

$$\mathcal{E} \otimes_{\mathcal{O}_{\text{Spec } A_k}} k(y) \longrightarrow H^0(C_{k(y)}, \mathcal{P}_y)$$

is an isomorphism.

We have the following sequence of morphisms:

$$C \times_k \text{Spec } A_k \xrightarrow{q_{A_k}} \mathbb{P}_{\times_k}^1 \text{Spec } A_k \xrightarrow{p_{A_k}} \text{Spec } A_k.$$

We continue to denote by π the projection onto the second factor of $C \times \text{Spec } A_k$ (so that $p_{A_k} \circ q_{A_k} = \pi$). Let \mathcal{F} denote the coherent sheaf $(q_{A_k})_* \mathcal{P}$ on $\mathbb{P}_k^1 \times_k \text{Spec } A_k$. The following lemma justifies this notation (compare Proposition 3.4).

Lemma 3.7. *For any closed point y in $\text{Spec } A_k$, let $i : \text{Spec } k(y) \rightarrow \text{Spec } A_k$ be the corresponding inclusion. Consider the Cartesian square*

$$\begin{array}{ccc} C \times_k \text{Spec } k(y) & \xrightarrow{\text{id} \times i} & C \times_k \text{Spec } A_k \\ q_{k(y)} \downarrow & & \downarrow q_{A_k} \\ \mathbb{P}_k^1 \times_k \text{Spec } k(y) & \xrightarrow{\text{id} \times i} & \mathbb{P}_k^1 \times_k \text{Spec } A_k \end{array}$$

Then the coherent sheaves $(\text{id} \times i)^(q_{A_k})_* \mathcal{P}$ and $(q_{k(y)})_*(\text{id} \times i)^* \mathcal{P}$ are isomorphic.*

Proof. Since \mathcal{P} is a coherent sheaf on the projective scheme $C \times_k \text{Spec } A_k$ (over $\text{Spec } A_k$), there exists a graded $(A_k[u, v, w]/(w^d - f))$ -module M such that \tilde{M} is isomorphic (as an $\mathcal{O}_{C \times_k \text{Spec } A_k}$ -module) to \mathcal{P} . Then the sheaf $(\text{id} \times i)^*(q_{k(y)})_* \mathcal{P}$ is isomorphic to $(k(y)_{[u, v]}(k(y) \otimes_{A_k} M))^\sim$. Also the sheaf $(q_{k(y)})_*(\text{id} \times i)^* \mathcal{P}$ is isomorphic to the sheaf $(k(y) \otimes_{A_k} A_k[u, v]M)^\sim$. These conclusions follow from [6] II, Proposition 2.8.10. But the graded $k(y)[u, v]$ -modules $(k(y)_{[u, v]}(k(y) \otimes_{A_k} A_k M))$ and $(k(y) \otimes_{A_k} A_k[u, v]M)$ are isomorphic. So the sheaves in question are isomorphic. \square

Our goal for this section is to find a graded $A_k[u, v]$ -module whose associated sheaf is isomorphic to \mathcal{F} . In fact, we will compute the module associated to \mathcal{F} , which will be sufficient. The next proposition is the main tool in this computation. Recall that for any morphism $g : X \rightarrow Y$ and a sheaf \mathcal{G} of \mathcal{O}_X -modules, there is a natural morphism $g^*g_*\mathcal{G} \rightarrow \mathcal{G}$. This follows from the adjointness of the functors g^* and g_* .

Proposition 3.8. *With the notation as before, the natural morphism*

$$u : p_{A_k}^*(p_{A_k})_*\mathcal{F} \longrightarrow \mathcal{F}$$

is an isomorphism.

Proof. Recall that $(p_{A_k})_*\mathcal{F} = \mathcal{E}$ is a locally free sheaf of rank d . So the sheaf $p_{A_k}^*(p_{A_k})_*\mathcal{F}$ is also a locally free sheaf of rank d on $\mathbb{P}_{A_k}^1$ ($\mathbb{P}_{A_k}^1 = \mathbb{P}_k^1 \times_k \operatorname{Spec} A_k$).

First we claim that \mathcal{F} is a locally free sheaf of rank d on $\mathbb{P}_{A_k}^1$. We prove that $\dim_{k(y)} \mathcal{F} \otimes_{\mathcal{O}_{\mathbb{P}_{A_k}^1}} k(y)$ is d for any closed point y in $\mathbb{P}_{A_k}^1$. Indeed by the upper semicontinuity of the dimension function and by the fact that $\mathbb{P}_{A_k}^1$ is of finite type over k , this will be sufficient.

Now for any closed point y in $\mathbb{P}_{A_k}^1$, consider the following diagram:

$$\begin{array}{ccc} & \operatorname{Spec} k(y) & \\ & \swarrow i \quad \downarrow j & \\ \mathbb{P}_{k(y)}^1 & \xrightarrow{\operatorname{id} \times i_y} & \mathbb{P}_{A_k}^1 \\ p_{k(y)} \downarrow & & \downarrow p_{A_k} \\ \operatorname{Spec} k(y) & \xrightarrow{i_y} & \operatorname{Spec} A_k \end{array}$$

By the universal property of the fibre product, the dotted arrow i exists so that this is a commutative diagram. We prove that

$$\dim_{k(y)} \mathcal{F} \otimes_{\mathcal{O}_{\mathbb{P}_{A_k}^1}} k(y) = \dim_{k(y)} j^* \mathcal{F}$$

is d . But

$$\dim_{k(y)} j^* \mathcal{F} = \dim_{k(y)} i^*(\operatorname{id} \times i_y)^* \mathcal{F}.$$

From Lemma 3.7 and Proposition 3.4, it follows that $(\operatorname{id} \times i_y)^* \mathcal{F}$ is a trivial vector bundle of rank d , so that $\dim_{k(y)} i^*(\operatorname{id} \times i_y)^* \mathcal{F}$ is d .

Now we go back to the morphism u defined earlier. Notice that it is a morphism of vector bundles of rank d . To prove that u is an isomorphism, it is sufficient to show that $u_x : (p_{A_k}^*(p_{A_k})_*\mathcal{F})_x \rightarrow \mathcal{F}_x$ is bijective for all points x in $\mathbb{P}_{A_k}^1$. By [6] I, Corollary 0.5.5.7, it is sufficient to prove that u_x is surjective at closed points. The part about closed points follows from the fact that the set $\{x \in \mathbb{P}_{A_k}^1 \mid u_x \text{ is surjective}\}$ is open in $\mathbb{P}_{A_k}^1$. Then by [6] I, Corollary 0.5.5.6, it is sufficient to prove that

$$u_y \otimes \operatorname{id} : (p_{A_k}^*(p_{A_k})_*\mathcal{F})_y / \mathfrak{m}_y(p_{A_k}^*(p_{A_k})_*\mathcal{F})_y \longrightarrow \mathcal{F}_y / \mathfrak{m}_y \mathcal{F}_y$$

is surjective for all closed points y in $\mathbb{P}_{A_k}^1$. But this homomorphism is surjective if and only if the morphism

$$j^*u : j^*p_{A_k}^*(p_{A_k})_*\mathcal{F} \longrightarrow j^*\mathcal{F}$$

is surjective, which is the same as

$$i^*(\mathrm{id} \times i_y)^*u : i^*(\mathrm{id} \times i_y)^*p_{A_k}^*(p_{A_k})_*\mathcal{F} \longrightarrow i^*(\mathrm{id} \times i_y)^*\mathcal{F}$$

being surjective. So it is sufficient to prove that

$$(\mathrm{id} \times i_y)^*u : (\mathrm{id} \times i_y)^*p_{A_k}^*(p_{A_k})_*\mathcal{F} \longrightarrow (\mathrm{id} \times i_y)^*\mathcal{F}$$

is an isomorphism. But we have the isomorphism

$$(\mathrm{id} \times i_y)^*p_{A_k}^*(p_{A_k})_*\mathcal{F} \cong p_{k(y)}^*i_y^*(p_{A_k})_*\mathcal{F}.$$

Note that $i_y^*(p_{A_k})_*\mathcal{F}$ is a trivial vector bundle of rank d , and so

$$p_{k(y)}^*i_y^*(p_{A_k})_*\mathcal{F} \cong p_{k(y)}^*\left(\bigoplus_d \mathcal{O}_{\mathrm{Spec} \, k(y)}\right) \cong \bigoplus_d \mathcal{O}_{\mathbb{P}^1_{k(y)}}.$$

This gives that $(\mathrm{id} \times i_y)^*p_{A_k}^*(p_{A_k})_*\mathcal{F}$ is a trivial vector bundle. Hence the morphism $(\mathrm{id} \times i_y)^*u$ will be an isomorphism if it is so on the global sections. But the morphism

$$p_{k(y)}^*i_y^*(p_{A_k})_*\mathcal{F} \longrightarrow (\mathrm{id} \times i_y)^*\mathcal{F}$$

on the global sections is an isomorphism if the natural morphism

$$(p_{A_k})_*\mathcal{F} \otimes_{\mathcal{O}_{\mathrm{Spec} \, A_k}} k(y) \longrightarrow H^0(\mathbb{P}^1_{k(y)}, \mathcal{F}_y)$$

is an isomorphism. By the earlier part (dimension computation) of the proof and [18], Corollary 2, p. 50, this is indeed the case. □

A useful application of this proposition is that it gives a convenient way to describe the graded module associated to the sheaf \mathcal{F} . Consider the sheaf $(p_{A_k})_*\mathcal{F} = \pi_*(\mathcal{P})$ on $\mathrm{Spec} \, A_k$. As we saw before, this is a locally free sheaf of rank d on $\mathrm{Spec} \, A_k$. So we can find a projective A_k -module P such that $\pi_*(\mathcal{P}) \cong \tilde{P}$.

Corollary 3.9. *The $A_k[u, v]$ -module $\bigoplus_i H^0(\mathbb{P}^1_{A_k}, \mathcal{F}(i))$ is (graded) isomorphic to the module $M = P \otimes_{A_k} A_k[u, v]$.*

Proof. It is sufficient to compute the graded module associated to the coherent sheaf $p_{A_k}^*(p_{A_k})_*\mathcal{F}$, which we denote by \mathcal{G} . But, by the projection formula,

$$(p_{A_k})_*(\mathcal{G}(i)) \cong (p_{A_k})_*(\mathcal{F}) \otimes (p_{A_k})_*(\mathcal{O}_{\mathbb{P}^1_{A_k}}(i)).$$

Since $\mathbb{P}^1_{A_k} \rightarrow \mathrm{Spec} \, A_k$ is a projective morphism, the module in the statement can be obtained by computing $H^0(\mathrm{Spec} \, A_k, (p_{A_k})_*(\mathcal{F}) \otimes (p_{A_k})_*(\mathcal{O}_{\mathbb{P}^1_{A_k}}(i)))$ for any i . From the above isomorphism it follows that

$$\bigoplus_i H^0(\mathrm{Spec} \, A_k, (p_{A_k})_*(\mathcal{F}) \otimes (p_{A_k})_*(\mathcal{O}_{\mathbb{P}^1_{A_k}}(i))) \cong P \otimes_{A_k} \left(\bigoplus_i (p_{A_k})_*(\mathcal{O}_{\mathbb{P}^1_{A_k}}(i)) \right).$$

But now the statement follows from the well-known computation of the graded module associated to the structure sheaf of the projective n -space. See, for example, [11], Chapter V, Proposition 5.13. □

4. THE HOMOMORPHISM $\tilde{C}_f \rightarrow \text{End}_{A_k}(P)$

We recall the set-up of the reduced Clifford algebra. Let f be a binary form of degree d over a field k such that $\text{char}(k)$ does not divide d . Then the Clifford algebra of the form f was defined in the introduction. See Section 1. One of the key properties of the representations of C_f is that the dimension of the representation is divisible by d ([10], Proposition 1.1). We are interested in representations of C_f of dimension d . We form the reduced Clifford algebra \tilde{C}_f by $(C_f)/(\bigcap \mathfrak{p})$, where the intersection is taken over the kernels of all dimension d representations. The algebra \tilde{C}_f is in fact an Azumaya algebra (for example, [10], Proposition 1.4), and so the d -dimensional representations are parametrized by the prime ideals in the center of \tilde{C}_f . Furthermore, it was pointed out to us that the center of \tilde{C}_f is Noetherian. This, for example, follows from Proposition 2 of [17], since C_f is finitely generated over k and \tilde{C}_f is a finite module over its center. We are trying to prove that the center of \tilde{C}_f is isomorphic to the k -algebra A_k of Section 3.

In this section, we consider the case when the curve C of Section 3 has a k -rational point. We construct a homomorphism from the center of \tilde{C}_f to A_k . This is achieved by first constructing a homomorphism from \tilde{C}_f to $\text{End}_{A_k}(P)$, where P is the module of the global sections of the sheaf $\pi_*\mathcal{P}$ on $\text{Spec } A_k$. See Section 3. We will show in the following sections that under this homomorphism the center of \tilde{C}_f is mapped to A_k . After we have obtained the desired homomorphism, we investigate some of its properties. In particular, we want to show how, starting with a finite-dimensional representation of \tilde{C}_f , we can construct a line bundle over C (possibly after a base extension).

Before we get to the main proposition of this section, we note an algebraic relation which holds inside the (graded) endomorphism ring of the graded module in Corollary 3.9. This relation will then enable us to define the morphism mentioned in the previous paragraph.

For the rest of this section, we assume that the curve C has a k -rational point.

Lemma 4.1. *Let P be the projective A_k -module as in Corollary 3.9. Consider the graded $A_k[u, v]$ -module $M = P \otimes_{A_k} A_k[u, v]$. Let P_i denote the i^{th} graded piece of this module (so $P \cong P_0$). Then we have the following equation:*

$$\text{Hom}_{A_k}(P_0, P_1) = u \circ \text{End}_{A_k} P_0 + v \circ \text{End}_{A_k} P_0,$$

where u, v are viewed in $\text{Hom}_{A_k}(P_0, P_1)$ with their natural action.

Proof. The proof is easy and follows from the relation $P_1 \cong (P_0 \oplus P_0)$, where the two generators are u and v . \square

Proposition 4.2. *There exists an algebra homomorphism*

$$\varphi : \tilde{C}_f \longrightarrow \text{End}_{A_k} P,$$

where \tilde{C}_f is the reduced Clifford algebra and P is as defined before.

Proof. The strategy to prove this proposition will be to show that $\text{End}_{A_k} P$ has elements that satisfy the relations of the Clifford algebra.

Consider the universal line bundle \mathcal{P} on $C \times_k A_k$ and the projection $q_{A_k} : C \times_k A_k \rightarrow \mathbb{P}_{A_k}^1$. Since q_{A_k} is a projective (and hence an affine) morphism, we can obtain the graded module associated to \mathcal{P} by computing $\bigoplus_i H^0(\mathbb{P}_{A_k}^1, (q_{A_k})_*(\mathcal{P}(i)))$. But this was already computed in Corollary 3.9. So the module M of Lemma 4.1 is

also a graded $A_k[u, v, w]/(w^d - f)$ -module. Since w is a homogeneous element of degree one in this (naturally) graded ring, it is an element of $\text{Hom}_{A_k}(P_0, P_1)$. By Lemma 4.1, there exist elements α_u, α_v in $\text{End}_{A_k} P$ such that the equality

$$w = u \circ \alpha_u + v \circ \alpha_v$$

holds in $\text{Hom}_{A_k}(P_0, P_1)$. Now consider the element w^d as an element of the module $\text{Hom}_{A_k}(P_0, P_d)$. Since in the graded ring of $C \times_k A_k$ we have the relation $w^d = f$, it holds in $\text{Hom}_{A_k}(P_0, P_d)$ as well. In particular, we get

$$w^d = (u \circ \alpha_u + v \circ \alpha_v)^d = f(u, v)\text{Id}.$$

But note that $\text{Hom}_{A_k}(P_0, P_d)$ is a free $\text{End}_{A_k} P_0$ -module. This shows that the elements α_u and α_v in $\text{End}_{A_k}(P_0)$ satisfy the relations of the Clifford algebra C_f and give a homomorphism

$$\tilde{\varphi} : C_f \longrightarrow \text{End}_{A_k} P_0.$$

Next we want to show that this homomorphism factors through the reduced Clifford algebra. Recall that $\text{Spec } A_k$ is an open subscheme in $\text{Pic}_{C/k}^{d+g-1}$, which is integral. Thus $\text{Spec } A_k$ is an integral scheme of finite type over k . In particular, A_k is an integral domain. So we may consider the sequence of morphisms

$$C_f \longrightarrow \text{End}_{A_k}(P_0) \longrightarrow \text{End}_{Q(A_k)}(P_0 \otimes_{A_k} Q(A_k)) \cong M_d(Q(A_k)),$$

where $Q(A_k)$ is the field of fractions of A_k . Note that, by the comments made above, the second arrow is an injection. This gives a d -dimensional representation of C_f . So the composite factors through \tilde{C}_f . We consider the diagram

$$\begin{array}{ccccc} C_f & \longrightarrow & \text{End}_{A_k} P_0 & \longrightarrow & M_d(Q(A_k)) \\ & \searrow & \uparrow & \nearrow & \\ & & \tilde{C}_f & & \end{array}$$

The dashed arrow exists since the second horizontal arrow is injective. This gives us the desired homomorphism

$$\varphi : \tilde{C}_f \longrightarrow \text{End}_{A_k} P_0.$$

□

The homomorphism $\tilde{C}_f \rightarrow \text{End}_{A_k} P_0$ constructed in the last proposition will always be denoted by φ .

Next we want to study the homomorphism φ . To this end, we assume that we are given a representation of \tilde{C}_f ,

$$\eta : \tilde{C}_f \longrightarrow M_d(K),$$

where K is a field extension of k . Our goal is to show that any such representation factors through $\text{End}_{A_k} P_0$ via φ .

We denote the images of the two generators of \tilde{C}_f under η by $\bar{\alpha}_u$ and $\bar{\alpha}_v$. Consider the morphism

$$\begin{aligned} S_K &= \frac{K[u, v, w]}{(w^d - f)} \longrightarrow M_d(K[u, v]), \\ u, v &\longrightarrow [u], [v], \\ w &\longrightarrow u\bar{\alpha}_u + v\bar{\alpha}_v. \end{aligned}$$

With the natural grading on $N = \bigoplus_d K[u, v]$, the above morphism is a graded homomorphism. This makes N a graded S_K -module. So \tilde{N} is a quasi-coherent (in fact, coherent) sheaf on $X = \text{Proj } S_K$. But we can do better, as is seen in the next lemma. Note that X is canonically isomorphic to $C_K = C \times_k K$.

Lemma 4.3. *\tilde{N} is an invertible sheaf on X . Moreover, the pull-back of this sheaf to C_K under the canonical morphism is a degree $(d + g - 1)$ line bundle on C_K . The image of the unique morphism $\text{Spec } K \rightarrow \text{Pic}_{C/k}^{d+g-1}$ corresponding to this line bundle lies in $\text{Spec } A_k$.*

Proof. For the first part of the statement, by [6], 4₂, Proposition 2.5.1, it is sufficient to prove that \tilde{N} is invertible with the assumption that K is algebraically closed. We will prove that for any closed point $x \in X$, $\dim_K(\tilde{N} \otimes_{\mathcal{O}_{X,x}} K) = 1$. This will be sufficient by the usual upper semicontinuity argument. Furthermore, it is clear that both u and v cannot be in a homogeneous maximal ideal of S_K . So it will be sufficient to prove the dimension condition for any closed point x in $X_v = \text{Spec}((S_K)_{(v)})$, since the argument for X_u is similar. Now

$$(S_K)_{(v)} \cong \frac{K[\bar{u}, \bar{w}]}{(\bar{w}^d - f(\bar{u}, 1))} \text{ and } N_{(v)} \cong \bigoplus_d K[\bar{u}],$$

where $\bar{u} = u/v$ and $\bar{w} = w/v$. Here \bar{u} acts in a natural way and \bar{w} acts as $\bar{u}\bar{\alpha}_u + \bar{\alpha}_v$. Any closed point in $\text{Spec}((S_K)_{(v)})$ is $\mathfrak{m} = (\bar{u} - a, \bar{w} - b)$ for some $a, b \in K$. So we have that

$$\mathcal{O}_{\text{Spec}((S_K)_{(v)}), x} \cong \left(\frac{K[\bar{u}, \bar{w}]}{(\bar{w}^d - f(\bar{u}, 1))} \right)_{(\bar{u}-a, \bar{w}-b)}$$

as well as

$$(N_{(v)})_x \cong \mathcal{O}_x \otimes_{(S_K)_{(v)}} \left(\bigoplus_d K[\bar{u}] \right).$$

This gives

$$\begin{aligned} (\tilde{N})_x \otimes_{\mathcal{O}_{X,x}} K &\cong \frac{\bigoplus_d K[\bar{u}]}{(\bar{u} - a, \bar{w} - b)(\bigoplus_d K[\bar{u}])} \\ &\cong \frac{\bigoplus_d K}{(a\bar{\alpha}_u + \bar{\alpha}_v - b)(\bigoplus_d K)}. \end{aligned}$$

So the required dimension is $\dim_K(\ker(a\bar{\alpha}_u + \bar{\alpha}_v - b))$. We consider $a\bar{\alpha}_u + \bar{\alpha}_v \in M_d(K)$ and compute the dimension of its eigenspace of eigenvalue b . The characteristic polynomial of $a\bar{\alpha}_u + \bar{\alpha}_v$ is $t^d - f(a, 1)$. Indeed, this follows when $f(a, 1) \neq 0$ since all the roots are distinct (since $\text{char}(K)$ does not divide d), and when $f(a, 1) = 0$ since the matrix is nilpotent in this case. Moreover, if $b \neq 0$, then $f(a, 1) \neq 0$ and b is an eigenvalue of multiplicity 1.

Let $b = 0$, so that $f(a, 1) = 0$. We may find a matrix $B \in \text{GL}_d(K)$ such that $B(a\bar{\alpha}_u + \bar{\alpha}_v)B^{-1}$ is in its Jordan form. If $\dim_K(\ker(a\bar{\alpha}_u + \bar{\alpha}_v - b)) > 1$, then we

can write (for $w \in M_d(K[u, v])$) $\det w = \det BwB^{-1} = (av - u)^l \det w'$, where

$$BwB^{-1} = \begin{pmatrix} av - u & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & av - u & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{pmatrix} w'.$$

There are l entries $(av - u)$ in the above matrix, and $l \geq 2$. But this gives that $\det w^d = (av - u)^{ld} \det(w')^d = f(u, v)^d$. Since $l \geq 2$, $av - u$ is a repeated factor of $f(u, v)$, which contradicts our assumption on f . This gives the required dimension condition, and \tilde{N} is an invertible sheaf on C_K .

For the second part, consider the projection $q_K : C_K \rightarrow \mathbb{P}_K^1$ considered before. Then

$$\chi(C_K, \tilde{N}) = \chi(\mathbb{P}_K^1, (q_K)_* \tilde{N}),$$

and, by the Riemann-Roch formula for vector bundles on nonsingular curves ([5], Example 15.2.1),

$$(1 - g) + \deg \tilde{N} = d(1 - 0) + \deg ((q_K)_* \tilde{N}).$$

But $(q_K)_* \tilde{N} \cong \bigoplus_d \mathcal{O}_{\mathbb{P}_K^1}$ and so $\deg ((q_K)_* \tilde{N}) = 0$. This gives that $\deg \tilde{N} = (d + g - 1)$.

For the third part, note that by the projection formula,

$$h^0(C_K, \tilde{N}(-1)) = h^0(\mathbb{P}_K^1, ((q_K)_* \tilde{N})(-1)) = 0.$$

By Proposition 2.6, the image of $\text{Spec } K$ under the unique morphism corresponding to \tilde{N} lies in $\text{Spec } A_k$. \square

Remark 4.4. The above lemma is essentially the key lemma of [22]. As we will see later, it provides an explicit construction of line bundles on $C \times_k K$ arising from representations of C_f . It is the main tool in proving the one-to-one correspondence between representations of \tilde{C}_f and line bundles on C mentioned in the introduction (Section 1).

Now we get to the main proposition mentioned at the beginning of the section.

Proposition 4.5. *Let $\eta : \tilde{C}_f \rightarrow M_d(K)$ be a representation. Then there exists a morphism $\text{End}_{A_k} P_0 \rightarrow M_d(K)$ so that the following diagram is commutative:*

$$\begin{array}{ccc} \tilde{C}_f & \xrightarrow{\varphi} & \text{End}_{A_k} P_0 \\ \eta \downarrow & \swarrow & \\ M_d(K) & & \end{array}$$

Proof. Given a finite-dimensional representation of \tilde{C}_f , we can construct a graded S_K -module N as before, where $S_K = (K[u, v, w]/(w^d - f))$. This gives a line bundle \mathcal{L} on C_K as in the last lemma. Then the image of the morphism corresponding to

this line bundle lies in $\text{Spec } A_k$. Note that by Lemma 2.13, this morphism does not depend on the choice of rigidification. If we denote by i the morphism

$$\text{Spec } K \xrightarrow{i} \text{Spec } A_k,$$

then $(\text{id} \times i)^*\mathcal{P} \cong \mathcal{L}$, where \mathcal{P} is the universal line bundle on $C \times_k \text{Spec } A_k$. Recall that \mathcal{P} is the sheaf associated to the graded module $\bigoplus_i P_i$, where $P_i = H^0(C \times_k \text{Spec } A_k, \mathcal{P}(i))$. In fact,

$$\bigoplus_i P_i = P \otimes_{A_k} A_k[u, v].$$

So the line bundle $(\text{id} \times i)^*\mathcal{P} \cong \mathcal{L} \cong \tilde{N}'$, where N' is the graded S_K -module $K \otimes_{A_k} (\bigoplus_i P_i)$. We claim that the S_K -modules N' and N are isomorphic as graded modules.

For any line bundle \mathcal{L}' , we denote by $\Gamma_*(\mathcal{L}')$ the associated graded module $\bigoplus_i H^0(\mathcal{L}'(i))$. We have the following diagram, in which the horizontal arrow exists by functoriality of Γ_* and the vertical morphisms are natural:

$$\begin{array}{ccc} \Gamma_*(\tilde{N}) & \longrightarrow & \Gamma_*(\tilde{N}') \\ \uparrow & & \uparrow \\ N & & N' \end{array}$$

Since the sheaves associated to the graded modules N, N' are isomorphic, the horizontal arrow is an isomorphism (as $\Gamma_*(\mathcal{O}_{C_K})$ and hence as S_K -modules). Thus it will be sufficient to prove that

- (1) $N \cong \Gamma_*(\tilde{N})$ and
- (2) $N' \cong \Gamma_*(\tilde{N}')$

as S_K -modules. Note that the vertical morphisms are S_K -module maps, and hence it is sufficient to prove that these are bijective. Consider the first morphism as a $K[u, v]$ -module morphism. By [11], Proposition 5.13 and the fact that Γ_* commutes with formation of direct sum, it follows that the first morphism is an isomorphism. But a similar argument along with the observation that

$$\begin{aligned} N' &\cong (K \otimes_{A_k} P_0) \otimes_{A_k} A_k[u, v] \\ &\cong \left(\bigoplus_d K\right) \otimes_{A_k} A_k[u, v] \\ &\cong \bigoplus_d K[u, v] \end{aligned}$$

shows that $N' \rightarrow \Gamma_*(\tilde{N}')$ is bijective.

This gives a degree-preserving S_K -module isomorphism

$$\bigoplus_d K[u, v] = N \xrightarrow{\mu} N' = (K \otimes_{A_k} P_0) \otimes_{A_k} A_k[u, v].$$

In particular, we get a K -module isomorphism

$$\bigoplus_d K \xrightarrow{\bar{\mu}} K \otimes_{A_k} P_0,$$

and hence an isomorphism

$$M_d(K) \xrightarrow{\bar{\mu}-\bar{\mu}^{-1}} \text{End}_K(K \otimes_{A_k} P_0).$$

Since μ preserves the action of w , the image of $\bar{\alpha}_u$ (respectively $\bar{\alpha}_v$) is $(\text{id} \otimes \alpha_u)$ (respectively $(\text{id} \otimes \alpha_v)$). So we have a commutative diagram

$$\begin{array}{ccc} \tilde{C}_f & \xrightarrow{\varphi} & \text{End}_{A_k} P_0 \\ \eta \downarrow & & \downarrow \\ M_d(K) & \xrightarrow{\cong} & \text{End}_K(K \otimes_{A_k} P_0) \end{array}$$

This is the desired diagram of the statement. \square

Remark 4.6. Note that the morphism $\text{End}_{A_k} P_0 \rightarrow M_d(K)$ obtained in the previous proposition by our procedure is uniquely determined up to a unique automorphism of $M_d(K)$. In fact, the only place where we do not have an explicit construction is in the use of the morphism $\text{Spec } K \rightarrow \text{Spec } A_k$. However, this morphism is uniquely determined once we have constructed a line bundle \tilde{N} as in the proof. This follows by Lemma 2.13. The rest of the procedure is algorithmic, and determines a unique homomorphism up to an element of $\text{GL}_d(K)$.

We now use the above proposition to prove a very useful property of the morphism φ .

Corollary 4.7. *The morphism*

$$\varphi : \tilde{C}_f \longrightarrow \text{End}_{A_k} P_0$$

constructed earlier is injective.

Proof. Let I denote the kernel of the morphism φ . Then since any representation has a factorization as in Proposition 4.5, I must be contained in the intersection of the kernels of d -dimensional representations. Thus, since \tilde{C}_f is an Azumaya algebra, I is contained in the (two-sided) ideal generated by the nilradical of Z_k , the center of \tilde{C}_f . But the center Z_k of \tilde{C}_f is reduced; see [10], Proposition 1.4. So the ideal $I = (0)$, and φ is injective. \square

4.1. The morphism $\varphi : \text{Spec } A_k \rightarrow \text{Spec } Z_k$. Consider the morphism $\varphi : \tilde{C}_f \rightarrow \text{End}_{A_k} P_0$ as above.

Proposition 4.8. *The morphism $\varphi : \tilde{C}_f \rightarrow \text{End}_{A_k} P_0$ maps the center Z_k to A_k .*

Proof. Consider the sequence of morphisms

$$\tilde{C}_f \xrightarrow{\varphi} \text{End}_{A_k} P_0 \xrightarrow{i} M_d(Q(A_k)).$$

Let $z \in Z_k$. We claim that $\varphi(z)$ is integral over A_k and that $i \circ \varphi(z) \in Q(A_k)$.

The first claim follows since $\varphi(z)$ commutes with A_k and $\text{End}_{A_k} P_0$ is a finitely generated module over A_k . For the second claim, note that the composite $i \circ \varphi$ is an irreducible representation of \tilde{C}_f . Hence it follows that $(i \circ \varphi)(\tilde{C}_f)Q(A_k) = M_d(Q(A_k))$. Now the claim follows since this implies that $i \circ \varphi(Z_k) \subseteq Q(A_k)$.

But since $\text{Spec } A_k$ is a smooth integral scheme, A_k is integrally closed, which implies that $\varphi(z) \in A_k$. \square

The above proposition along with Corollary 4.7 gives an injective k -algebra morphism

$$\varphi : Z_k \longrightarrow A_k$$

and hence a k -morphism

$$\varphi : \operatorname{Spec} A_k \longrightarrow \operatorname{Spec} Z_k.$$

If there is no cause for confusion, we will denote both of the above morphisms by φ .

We will prove in the following sections that φ is an isomorphism.

5. THE MORPHISM $\eta : \operatorname{Spec} A_k \rightarrow \operatorname{Spec} Z_k$

In this section, we construct a morphism $\eta : \operatorname{Spec} Z_k \rightarrow \operatorname{Spec} A_k$ which we will show to be the inverse of the morphism φ constructed earlier. The main idea in constructing the morphism η is to view the schemes $\operatorname{Spec} A_k$ and $\operatorname{Spec} Z_k$ as representing objects of certain functors, and then construct a natural transformation between these functors. These ideas are present in the paper [23] of M. Van den Bergh.

If S is a k -scheme, then by an S -representation of degree n of \tilde{C}_f we mean a pair (ψ, \mathcal{O}_A) , where \mathcal{O}_A is a sheaf of Azumaya algebras of rank n^2 over S and $\psi : \tilde{C}_f \rightarrow H^0(S, \mathcal{O}_A)$ is a k -algebra homomorphism. Two S -representations $(\psi_1, \mathcal{O}_{A_1})$ and $(\psi_2, \mathcal{O}_{A_2})$ are said to be equivalent if there is an isomorphism $\theta : \mathcal{O}_{A_1} \rightarrow \mathcal{O}_{A_2}$ of sheaves of Azumaya algebras such that $\psi_2 = H^0(S, \theta) \circ \psi_1$. An S -representation of \tilde{C}_f is irreducible if the image of \tilde{C}_f generates \mathcal{O}_A locally.

Let $\operatorname{Rep}_d(\tilde{C}_f, -)$ be the functor that assigns to a k -scheme S the set of equivalence classes of irreducible S -representations of degree d of \tilde{C}_f . Since Azumaya algebras pull back to Azumaya algebras and irreducible representations are stable under pull-back, it follows that this is a functor. Furthermore, since \tilde{C}_f is an Azumaya algebra of rank d^2 over its center, this functor is representable in Sch/k and is represented by $\operatorname{Spec} Z_k$. Note that, by remarks at the beginning of Section 4, $\operatorname{Spec} Z_k$ is a Noetherian scheme.

We also consider representations of \tilde{C}_f into endomorphism sheaves of vector bundles. Let $\mathcal{G}_d(\tilde{C}_f, -)$ be the subfunctor of $\operatorname{Rep}_d(\tilde{C}_f, -)$ that assigns to a k -scheme S the set of equivalence classes of irreducible S -representations into endomorphism sheaves of vector bundles of rank d . Again, since endomorphism sheaves of vector bundles pull back to the sheaves of the same kind, it follows that this is also a functor. Both the functors $\operatorname{Rep}_d(\tilde{C}_f, -)$ and $\mathcal{G}_d(\tilde{C}_f, -)$ define presheaves on $(\operatorname{Sch}/k)_{\text{fl}}$ with respect to the flat topology. Moreover, the functor $\operatorname{Rep}_d(\tilde{C}_f, -)$ is a sheaf with respect to the flat topology, since it is representable (Proposition 1, Chapter 8, [2]). We denote the sheafification of the functor $\mathcal{G}_d(\tilde{C}_f, -)$ with respect to the flat topology by $\bar{\mathcal{G}}_d(\tilde{C}_f, -)$. The following fact is [23], Lemma 4.2.

Lemma 5.1. *The natural transformation $\mathcal{G}_d(\tilde{C}_f, -) \rightarrow \operatorname{Rep}_d(\tilde{C}_f, -)$ induces an isomorphism $\operatorname{Rep}_d(\tilde{C}_f, -) \cong \bar{\mathcal{G}}_d(\tilde{C}_f, -)$.*

Our next goal is to define a natural transformation from $\operatorname{Rep}_d(\tilde{C}_f, -)$ to a certain open subfunctor of $\operatorname{Pic}_{C/k}^{d+g-1}(-)$. Before we define this morphism of functors, we need some auxiliary lemmas.

Let S be a k -scheme, and let (ψ, \mathcal{O}_A) be an element of $\mathcal{G}_d(\tilde{C}_f, S)$ so that $\mathcal{O}_A = \mathcal{E}nd_{\mathcal{O}_S}(\mathcal{E})$ for some vector bundle \mathcal{E} of rank d on $C \times_k S$. We will construct an invertible sheaf on $C \times_k S$. Consider the graded sheaf homomorphism

$$\begin{aligned} \frac{\mathcal{O}_S[u, v, w]}{w^d - f(u, v)} &\longrightarrow \mathcal{E}nd(\mathcal{E})[u, v], \\ u, v &\longrightarrow u, v, \\ w &\longrightarrow u\psi(x_1) + v\psi(x_2), \end{aligned}$$

where x_1 and x_2 are the usual generators of \tilde{C}_f . Since \mathcal{E} is a sheaf of \mathcal{O}_S -modules, there is a canonical map from \mathcal{O}_S to $\mathcal{E}nd(\mathcal{E})$. Since this is a graded homomorphism of graded algebras (degree of u, v is 1 on the right side), we get a graded module $\mathcal{E} \otimes_{\mathcal{O}_S} \mathcal{O}_S[u, v]$. This defines a sheaf \mathcal{M} on $C \times_k S$. In the lemma below and in the remainder of this section, for any point s in S , q_s denotes the morphism $C \times_k k(s) \rightarrow \mathbb{P}_{k(s)}^1$ induced by the inclusion $k(s)[u, v] \rightarrow (k(s)[u, v, w])/(w^d - f(u, v))$, p_s denotes the projection of the second factor of $\mathbb{P}_{k(s)}^1$ onto $\text{Spec } k(s)$, and π_s denotes the composition $p_s \circ q_s$. Similar notation is used for an arbitrary field K instead of $k(s)$.

Lemma 5.2. *The sheaf \mathcal{M} is an invertible sheaf on $C \times_k S$ of (fibre-wise constant) degree $(d + g - 1)$. Furthermore, the invertible sheaf $\mathcal{M}_s = \mathcal{M} \otimes_{\mathcal{O}_S} \text{Spec } k(s)$ on $C_{k(s)}$ satisfies $h^0(C_{k(s)}, \mathcal{M}_s(-1)) = 0$.*

Proof. It is sufficient to prove the first assertion in the case when S is an affine scheme. So let $S = \text{Spec } R$. Let C_R denote the scheme $C \times_k S$. We may assume that $H^0(S, \mathcal{E}) \cong \bigoplus_d R$, since \mathcal{E} is a vector bundle of rank d . In this case, we denote the graded $(R[u, v, w]/(w^d - f(u, v))$ -module $H^0(S, \mathcal{E}) \otimes_R R[u, v]$ by M and the associated sheaf \mathcal{M} by \tilde{M} .

Now consider the first assertion. Note that \tilde{M} is flat over S and π is a flat morphism. So by [6], 4₂, Lemma 12.3.1, it is sufficient to prove that for any point s in S , \tilde{M}_s is an invertible sheaf on S . Here \tilde{M}_s is the sheaf $(1 \times i_s)^* \tilde{M}$ where $1 \times i_s$ is the morphism $C_s = C \times_k \text{Spec } k(s) \rightarrow C \times_k S$. Consider the representation associated to the point s via ψ :

$$\psi_s : \tilde{C}_f \longrightarrow M_d(k(s)).$$

This gives a sheaf \tilde{N} as in the discussion preceding Lemma 4.3. The sheaf \tilde{N} is isomorphic to the sheaf \tilde{M}_s . So the fact that \tilde{M}_s is invertible now follows from Lemma 4.3. Also, the degree is $(d + g - 1)$ along fibres of the projection onto the second factor. This follows from a calculation using the Riemann-Roch formula; see the proof of Lemma 4.3.

For the second assertion, it is sufficient to prove the statement for the invertible sheaf \tilde{N} of the last paragraph. But, in this case, $(q_{k(s)})_* \mathcal{M}_s$ is a trivial vector bundle of rank d . This, along with the projection formula, gives the required statement. See Lemma 4.3 for details. \square

Corollary 5.3. *With the notation as in Lemma 5.2, the resulting morphism $S \rightarrow \text{Pic}_{C/k}^{d+g-1}$ factors through $\text{Spec } A_k$.*

Proof. Since the sheaf \mathcal{M} is invertible on $C \times_k S$, we have a morphism $S \rightarrow \text{Pic}_{C/k}^{d+g-1}$ (after rigidifying this invertible sheaf if necessary). From Remark 2.15, this morphism does not depend on the rigidification. Now it is sufficient to prove that the

image of any point s in S is in $\operatorname{Spec} A_k$ under the morphism in the statement. However, it follows from Lemma 4.3 that the image of any point s in S lies in $\operatorname{Spec} A_k$. So we get the required factorization. \square

Now we consider another functor $\mathcal{H}_{C/k}^{d+g-1}$ defined as follows. This will be defined as a subfunctor of $\operatorname{Pic}_{C/k}^{d+g-1}$. For any k -scheme S , let $\mathcal{H}_{C/k}^{d+g-1}(S)$ be the subset of $\operatorname{Pic}_{C/k}^{d+g-1}(S)$ associated to the set of rigidified line bundles \mathcal{L} on $C \times_k S$ that satisfy the condition that for any point s of S , $h^0(C_s, \mathcal{L}_s(-1)) = 0$. Here $C_s = C \times_k k(s)$ and \mathcal{L}_s denotes the sheaf $(1_C \times i(s))^* \mathcal{L}$, where $i(s)$ is the inclusion $\operatorname{Spec} k(s) \rightarrow S$.

To see that this is indeed a functor, consider a k -scheme T and a k -morphism $f : T \rightarrow S$, and let \mathcal{L} be an invertible sheaf on $C_S = C \times_k S$ which gives an element of $\mathcal{H}_{C/k}^{d+g-1}(S)$. We need to prove that for any point t in T , $h^0(C_t, \mathcal{L}_t(-1)) = 0$. But if s is the image of t , then we have $h^0(C_s, \mathcal{L}_s(-1)) = 0$. Now the required dimension is zero, since the composition $f \circ i(t)$ factors through the inclusion $i(s) : \operatorname{Spec} k(s) \rightarrow S$. It follows that this definition gives a functor. For this functor, we have the following lemma.

Lemma 5.4. *The functor $\mathcal{H}_{C/k}^{d+g-1}$ is an open subfunctor of the functor $\operatorname{Pic}_{C/k}^{d+g-1}$. Furthermore, $\mathcal{H}_{C/k}^{d+g-1}$ is a representable functor and is represented by $\operatorname{Spec} A_k$.*

Proof. To verify the openness of $\mathcal{H}_{C/k}^{d+g-1}$, we use the characterization of openness of subfunctors given in [23], Lemma 2.2.3. For a k -scheme S , let \mathcal{L} be an invertible sheaf on $C \times_k S$. Then condition (2a) is satisfied by the definition of the functor. For condition (2b), let L/K be a field extension, and let $i : \operatorname{Spec} K \rightarrow \operatorname{Spec} L$ be the associated morphism. Since i is flat, it follows that $(\pi_L)_*(1_{C_k} \times i)^* \mathcal{L} \cong i^*(\pi_K)_* \mathcal{L}$. Now the condition follows since $((1_{C_k} \times i)^* \mathcal{L})(-1) \cong (1_{C_k} \times i)^*(\mathcal{L}(-1))$. For condition (2c), we need to prove that the set of points $\{s\}$ in S for which \mathcal{L}_s is in $\mathcal{H}_{C/k}^{d+g-1}(\operatorname{Spec} k(s))$ is open in S . We note that the image of a point s of S in $\operatorname{Pic}_{C/k}^{d+g-1}$ lies in $\operatorname{Spec} A_k$ if and only if $h^0(C_s, \mathcal{L}_s(-1)) = 0$ (Proposition 2.6). Now the condition follows by noting that the set in question is the inverse image of $\operatorname{Spec} A_k$ under the morphism $S \rightarrow \operatorname{Pic}_{C/k}^{d+g-1}$.

Moreover, by the discussion above, it also follows that the canonical morphism $S \rightarrow \operatorname{Pic}_{C/k}^{d+g-1}$ associated with an element of $\mathcal{H}_{C/k}^{d+g-1}(S)$ factors through $\operatorname{Spec} A_k$. Note that the restriction of the universal invertible sheaf to $C \times_k \operatorname{Spec} A_k$ is an element of $\mathcal{H}_{C/k}^{d+g-1}(\operatorname{Spec} A_k)$. This follows from Proposition 2.6. This shows that $\operatorname{Spec} A_k$ represents the functor $\mathcal{H}_{C/k}^{d+g-1}$. \square

We can now construct a natural transformation

$$\Phi : \mathcal{G}_d(\tilde{C}_f, -) \longrightarrow \mathcal{H}_{C/k}^{d+g-1}(-).$$

Namely, to any element $(\mathcal{E}nd(\mathcal{E}), \psi)$ in $\mathcal{G}_d(\tilde{C}_f, S)$, we assign the $\mathcal{O}_{C \times_k S}$ -module $(H^0(S, \mathcal{E}) \otimes_{\mathcal{O}_S} \mathcal{O}_S[u, v])^\sim$. To see that Φ is a natural transformation, let $f : T \rightarrow S$ be a morphism of schemes and let $(\mathcal{E}nd(\mathcal{E}), \psi)$ be an element of $\mathcal{G}_d(\tilde{C}_f, S)$. Then we need to prove that the two invertible $\mathcal{O}_{C \times_k T}$ -modules $\mathcal{O}_T \otimes_{\mathcal{O}_S} p_S^* \mathcal{E}$ and $p_T^* f^* \mathcal{E}$ are isomorphic. Now the statement is immediate, since the action of w is preserved by the natural $\mathcal{O}_{\mathbb{P}^1_T}$ -module isomorphism between these sheaves.

Since the functor $\mathcal{H}_{C/k}^{d+g-1}$ is representable, by sheafification of Φ , we get the natural transformation

$$\Phi : \mathcal{R}ep_d(\tilde{C}_f, -) \cong \bar{\mathcal{G}}_d(\tilde{C}_f, -) \longrightarrow \mathcal{H}_{C/k}^{d+g-1}(-).$$

The isomorphism in the above definition follows from Lemma 5.1. Now the functors $\mathcal{R}ep_d(\tilde{C}_f, -)$ and $\mathcal{H}_{C/k}^{d+g-1}(-)$ are represented by $\mathcal{S}pec Z_k$ and $\mathcal{S}pec A_k$, respectively. This follows from the fact that \tilde{C}_f is an Azumaya algebra over its center and Lemma 5.4. So the natural transformation induces a morphism of schemes

$$\eta : \mathcal{S}pec Z_k \longrightarrow \mathcal{S}pec A_k.$$

In the following section, we will show that the morphisms η and φ are inverses of each other.

6. THE ISOMORPHISM THEOREM AND DESCENT ON φ

In this section, we first prove that the morphism φ is an isomorphism. Now consider the composition

$$\mathcal{S}pec A_k \xrightarrow{\varphi} \mathcal{S}pec Z_k \xrightarrow{\eta} \mathcal{S}pec A_k,$$

and denote it by θ . In the following proposition, \mathcal{P} denotes the restriction of the universal invertible sheaf to $C \times_k \mathcal{S}pec A_k$.

Proposition 6.1. *The morphism θ is the identity morphism on $\mathcal{S}pec A_k$.*

Proof. Note that since \mathcal{P} is the universal object in $\mathcal{H}_{C/k}^{d+g-1}(\mathcal{S}pec A_k)$, it will be sufficient to prove that $(1_C \times \theta)^*\mathcal{P}$ is isomorphic to \mathcal{P} as a rigidified line bundle. Furthermore, by Remark 2.15 it is sufficient to show that these two sheaves are isomorphic as line bundles. Note that the sheaf $(1_C \times \theta)^*\mathcal{P}$ is isomorphic to the image under Φ of $(\varphi, \mathcal{E}nd(\mathcal{P}))$ in $\mathcal{R}ep_d(\tilde{C}_f, \mathcal{S}pec A_k)$. But then the sheaf $(1_C \times \theta)^*\mathcal{P}$ is isomorphic to $(P \otimes_{A_k} A_k[u, v])^\sim$, where $P = H^0(\mathcal{S}pec A_k, (\pi_{A_k})_*(\mathcal{P}))$ and w acts as $u\varphi(x_1) + v\varphi(x_2)$. Now from the construction of the morphism φ (Proposition 4.2), the two actions of w on the graded $(A_k[u, v, w]/(w^d - f(u, v)))$ -module $P \otimes_{A_k} A_k[u, v]$ corresponding to the invertible sheaves $(1_C \times \theta)^*\mathcal{P}$ and \mathcal{P} on $C \times_k \mathcal{S}pec A_k$ are compatible. So the sheaves in question are isomorphic. This shows that θ is the identity morphism. \square

Now we can prove the following theorem.

Theorem 6.2. *Let k be a field so that the curve C has a k -rational point. Then the morphism φ is an isomorphism.*

Proof. We prove that the corresponding homomorphism $\varphi : Z_k \rightarrow A_k$ is an isomorphism. By Proposition 4.8, it is sufficient to show that the map φ is surjective. By Proposition 6.1, there exists a ring homomorphism $\eta : A_k \rightarrow Z_k$ so that the composition

$$A_k \xrightarrow{\eta} Z_k \xrightarrow{\varphi} A_k$$

is the identity homomorphism. In particular, φ is surjective. It follows that φ is an isomorphism. \square

6.1. The Galois descent on φ . In this section, we want to prove the isomorphism theorem of the last section without the assumption that the curve C has a k -rational point. The idea is to prove that the φ obtained earlier descends to a field over which f is defined.

Let f be a binary form over a field k such that the characteristic of k does not divide d and f does not have repeated factors over an algebraic closure of k . Let k'/k be a finite Galois extension (with Galois group G) such that $C(k')$ is non-empty. Here we may assume that k' is a Galois extension of k by the assumption on $\text{char}(k)$. Then we have the isomorphism

$$\varphi : \text{Spec } A_{k'} \longrightarrow \text{Spec } Z_{k'}.$$

Recall that $\text{Spec } A_{k'}$ descends to $\text{Spec } A_k$, which is the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$. Also, $\text{Spec } Z_{k'}$ descends to $\text{Spec } Z_k$, which is the center of the Clifford algebra of f over k .

Main Theorem. *The morphism φ descends to k , and hence the center of the reduced Clifford algebra is isomorphic to the coordinate ring of the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$.*

Proof. Let x be a closed point of $\text{Spec } A_{k'}$, and let $k(x)$ be its residue field. Let $\sigma \in G$. Then we have the following diagram:

$$\begin{array}{ccc} \text{Spec } k(x) & & \\ \downarrow i & & \\ \text{Spec } A_{k'} & \xrightarrow{\varphi} & \text{Spec } Z_{k'} \\ \downarrow \sigma & & \downarrow \sigma \\ \text{Spec } A_{k'} & \xrightarrow{\varphi} & \text{Spec } Z_{k'} \end{array}$$

Now we prove that the above diagram is commutative. Let j denote the open immersion $\text{Spec } A_{k'} \hookrightarrow \text{Pic}_{C'/k'}^{d+g-1}$. We first show that

$$(6.1) \qquad j \circ \sigma \circ i = j \circ \varphi^{-1} \circ \sigma \circ \varphi \circ i.$$

Let π denote the structure morphism $\text{Pic}_{C'/k'}^{d+g-1} \rightarrow \text{Spec } k'$, and $C' \times_{k'} k(x)^\sigma$ the fibred product of C' and $\text{Spec } k(x)$ with the structure morphism for $\text{Spec } k(x)$ as $\pi \circ \sigma \circ j \circ i$. Also, let $C' \times_{k'} k(x)$ denote the fibred product as earlier with the structure morphism for $\text{Spec } k(x)$ as $\pi \circ j \circ i$. We have the isomorphism

$$(6.2) \qquad C' \times_{k'} k(x) \xrightarrow{\sigma \times \text{id}} C' \times_{k'} k(x)^\sigma.$$

To prove (6.1), it is sufficient to prove that the line bundles on $C' \times_{k'} k(x)^\sigma$ corresponding to these morphisms are isomorphic. See Lemma 2.13. If \mathcal{P} denotes the universal line bundle on $C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}$, then the line bundle corresponding to the morphism on the left side of (6.1) is $(\text{id} \times (j \circ \sigma \circ i))^* \mathcal{P}$.

Consider the morphism on the right side of the above equation. To construct the corresponding line bundle, it is sufficient to consider the associated representation of the Clifford algebra. The required line bundle corresponds to the following composite morphism:

$$\tilde{C}_f \otimes_k k' \xrightarrow{\text{id} \times \sigma} \tilde{C}_f \otimes_k k' \xrightarrow{\eta_i} M_d(k(x)),$$

where η_i is the representation corresponding to the morphism i (or $i \circ \varphi$). But this line bundle is isomorphic to the pull-back of \mathcal{P} under the composite morphism

$$C' \times_{k'} k(x)^\sigma \xrightarrow{\sigma^{-1} \times \text{id}} C' \times_{k'} k(x) \xrightarrow{\text{id} \times (j \circ i)} C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1}.$$

Now consider the commutative diagram

$$\begin{array}{ccc} C' \times_{k'} k(x) & \xrightarrow{\sigma \times \text{id}} & C' \times_{k'} k(x)^\sigma \\ \text{id} \times (j \circ i) \downarrow & & \downarrow \text{id} \times (j \circ \sigma \circ i) \\ C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1} & \xrightarrow{\sigma \times \sigma} & C' \times_{k'} \text{Pic}_{C'/k'}^{d+g-1} \end{array}$$

By the above argument and the commutativity of this diagram, it follows that the two line bundles in question are isomorphic to

$$(\text{id} \times (j \circ \sigma \circ i))^* \mathcal{P} \quad \text{and} \quad (\text{id} \times (j \circ \sigma \circ i))^* (\sigma^{-1} \times \sigma^{-1})^* \mathcal{P}.$$

By Lemma 2.12, these two line bundles are isomorphic. So (6.1) holds true.

In particular, $i = (\sigma^{-1} \circ \varphi^{-1} \circ \sigma \circ \varphi) \circ i$. Now consider the two automorphisms id and $\mu = (\sigma^{-1} \circ \varphi^{-1} \circ \sigma \circ \varphi)$ of $\text{Spec } A_{k'}$. They both are the identity on closed points, and the morphisms induced on the residue fields are the same. Let $a \in A_{k'}$. Then $(\mu(a) - a)$ is in any maximal ideal on $A_{k'}$. Since $A_{k'}$ is of finite type over k' and is an integral domain, the intersection of all the maximal ideals is the zero ideal. So $\mu(a) = a$ for any $a \in A_{k'}$. This shows that μ is the identity morphism.

The second part follows from the discussion preceding the theorem and Corollary 2.7. \square

We record an easy corollary of the main theorem.

Corollary 6.3. *The center Z_k of the reduced Clifford algebra is an integrally closed Noetherian domain.*

Proof. By the main theorem, the k -scheme $\text{Spec } Z_k$ is isomorphic to the scheme $\text{Spec } A_k$, which is the complement of the Θ -divisor in $\text{Pic}_{C/k}^{d+g-1}$. Since $\text{Pic}_{C/k}^{d+g-1}$ is a smooth, integral Noetherian scheme and $\text{Spec } A_k$ is an open subscheme of $\text{Pic}_{C/k}^{d+g-1}$, $\text{Spec } A_k$ has these properties as well. In particular, A_k is an integrally closed Noetherian domain. So Z_k is an integrally closed Noetherian domain. \square

Remark 6.4. The Main Theorem 6.1 says that the reduced Clifford algebra can be viewed as a sheaf of Azumaya algebras on $\text{Spec } A_k$. There are two natural questions which arise. The first is, does this sheaf extend to $\text{Pic}_{C/k}^{d+g-1}$ as an Azumaya algebra? The second is if it does, then what is its Brauer class? These questions will be addressed in a forthcoming article [13].

ACKNOWLEDGEMENTS

We thank Darrell Haile, Michael Larsen and Valery Lunts for their help and encouragement. This project has benefitted immensely from our conversations with them and from their advice.

REFERENCES

1. E. Arbarello, M. Cornalba, P. A. Griffiths and J. Harris, *Geometry of Algebraic Curves*, vol. 1, Springer-Verlag, New York, 1985. MR **86h**:14019
2. S. Bosch, W. Lütkebohmert and M. Raynaud, *Néron Models*, Springer-Verlag, New York, 1990. MR **91i**:14034
3. L. Childs, Linearizing of n -ic forms and generalized Clifford algebras, *Linear and Multilinear Algebra* **5** (1978), 267-278. MR **57**:12567
4. F. DeMeyer and E. Ingraham, *Separable Algebras over Commutative Rings, Lecture Notes in Math.*, vol. 181, Springer-Verlag, Berlin, 1971. MR **43**:6199
5. W. Fulton, *Intersection Theory*, Springer-Verlag, New York, 1998. MR **99d**:14003
6. A. Grothendieck and J. Dieudonné, *Eléments de Géométrie Algébrique, Inst. Hautes Études Sci. Publ. Math.*, nos. 4, 8, 11, 17, 20, 24, 28, 32 (1964-1967). MR **29**:1210; MR **30**:3885; MR **33**:7330; MR **36**:178; MR **39**:220
7. A. Grothendieck, *Technique de descente et théorèmes d'existence en géométrie algébrique. II: Le théorème d'existence en théorie formelle des modules, Séminaire Bourbaki 1959/60, Exposé 195*, Secrétariat Math., Paris, 1960 (and later reprints by other publishers). MR **23**:A2273
8. D. Haile, On the Clifford algebra of a binary cubic form, *Amer. J. Math.* **106** (1984), 1269-1280. MR **86c**:11028
9. D. Haile, When is the Clifford algebra of a binary cubic form split, *J. Algebra* **146** (1992), 514-520. MR **93a**:11029
10. D. Haile and S. Tesser, On Azumaya algebras arising from Clifford algebras, *J. Algebra* **116** (1988), 372-384. MR **89j**:15044
11. R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, 1977. MR **57**:3116
12. R. S. Kulkarni, On the Clifford algebra of a binary form, Ph.D. Thesis, Indiana University, 1999.
13. R. S. Kulkarni, On the extension of the Brauer class of the reduced Clifford algebra, submitted.
14. J. Milne, Abelian Varieties: *Arithmetic Geometry*, Springer-Verlag, 1986, pp. 103-150. MR **89b**:14029
15. J. Milne, Jacobian Varieties: *Arithmetic Geometry*, Springer-Verlag, 1986, pp. 167-212. MR **89b**:14029
16. J. Milne, *Étale Cohomology*, Princeton University Press, 1980. MR **81j**:14002
17. S. Montgomery and L. W. Small, Fixed rings of Noetherian rings, *Bull. London Math. Soc.* **13** (1981), 33-38. MR **82a**:16033
18. D. Mumford, *Abelian Varieties*, 2nd ed., Oxford University Press, 1970, MR **44**:219
19. C. Processi, *Rings with Polynomial Identities*, Marcel Dekker, New York, 1973. MR **51**:3214
20. P. Revoy, Algèbres de Clifford et algèbres extérieures, *J. Algebra* **46** (1977), 268-277. MR **57**:12568
21. N. Roby, Algèbres de Clifford des formes polynômes, *C. R. Acad. Sci. Paris Sér. I Math. A-B* **268** (1969), A484-A486. MR **39**:2794
22. M. Van den Bergh, Linearisations of binary and ternary forms, *J. Algebra* **109** (1987), 172-183. MR **88j**:11020
23. M. Van den Bergh, The center of the generic division algebra, *J. Algebra* **127** (1989), 106-126. MR **91d**:16031

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN-MADISON, MADISON, WISCONSIN 53706

Current address: Department of Mathematics, Wells Hall, Michigan State University, East Lansing, Michigan 48824

E-mail address: kulkarni@math.msu.edu

PROJECTIVE NORMALITY OF ABELIAN VARIETIES

JAYA N. IYER

ABSTRACT. We show that ample line bundles L on a g -dimensional simple abelian variety A , satisfying $h^0(A, L) > 2^g \cdot g!$, give projective normal embeddings, for all $g \geq 1$.

1. INTRODUCTION

Let A be an abelian variety of dimension g defined over the field of complex numbers and let L be an ample line bundle on A . Consider the associated rational map $\phi_L : A \rightarrow \mathbb{P}^{d-1} = \mathbb{P}H^0(L)$, where $d = \dim H^0(A, L)$. Suppose $L = M^n$ for some ample line bundle M on A . Then Koizumi has shown that L gives a projectively normal embedding if $n \geq 3$ (see [2]).

When $n = 2$, Ohbuchi (see [7]) has shown the following.

Theorem 1.1. *Suppose M is a symmetric ample line bundle on a g -dimensional abelian variety A . Then $L = M^2$ gives a projectively normal embedding of A if and only if the origin 0 of A is not contained in $Bs|M \otimes P_\alpha|$ for any $\alpha \in \hat{A}_2 = \{\alpha \in \hat{A} : 2\alpha = 0\}$, where \hat{A} is the dual abelian variety of A , P is the Poincaré bundle on $A \times \hat{A}$, $P_\alpha = P|_{A \times \alpha}$ for $\alpha \in \hat{A}$ and $Bs|M \otimes P_\alpha|$ is the set of all base points of $M \otimes P_\alpha$.*

Suppose $L \neq M^n$ for any ample line bundle M on A and $n > 1$. When $g = 2$, Lazarsfeld (see [4]) has shown that if ϕ_L is birational onto its image, then ϕ_L gives a projectively normal embedding, for $d = 7, 9, 11$ and for $d \geq 13$. We showed that if the Neron Severi group $NS(A)$ of A is \mathbb{Z} , generated by L and $d \geq 7$, then ϕ_L gives a projectively normal embedding (see [1]).

In this article, we show

Theorem 1.2. *Suppose L is an ample line bundle on a g -dimensional simple abelian variety A . If $d > 2^g \cdot g!$, then L gives a projectively normal embedding, for all $g \geq 1$. (Here $d = \dim H^0(A, L)$).*

We outline the proof of Theorem 1.2.

For a polarized abelian variety (A, L) , consider the multiplication maps

$$\rho_r : \text{Sym}^r H^0(A, L) \rightarrow H^0(A, L^r).$$

By definition, L gives a projectively normal embedding if ρ_r is surjective, for all $r \geq 1$. We first show that it suffices to show ρ_2 is surjective. More precisely, we show that ρ_2 surjective implies that the maps ρ_r are surjective, for $r \geq 3$ (see Prop. 2.1).

Received by the editors December 5, 2001 and, in revised form, October 20, 2002.
 2000 *Mathematics Subject Classification*. Primary 14C20, 14K05, 14K25, 14N05.

To prove the surjectivity of the map ρ_2 we consider a finite isogeny $A \longrightarrow B = A/H$, where H is a maximal isotropic subgroup of the fixed group $K(L)$ of L . Then L descends down to a principal polarization M on B . Let \hat{H} denote the group of characters on H . By associating to a character $\chi \in \hat{H}$ a degree 0 line bundle L_χ on B one can identify \hat{H} as a subgroup of the dual abelian variety $\text{Pic}^0(B)$ of B . The homomorphism $\psi_M : B \longrightarrow \text{Pic}^0(B), b \mapsto t_b^*M \otimes M^{-1}$ is an isomorphism and we denote $H' = \psi_M^{-1}(\hat{H})$.

We then show that the surjectivity of the map ρ_2 is equivalent to showing that the subgroup H' of B generates the projective space $\mathbb{P}H^0(B, M^2)$ and its translates $\mathbb{P}H^0(t_\sigma^*M^2)$, where $\sigma \in B$ is such that $\psi_M(2\sigma) = L_\chi, L_\chi \in \hat{H}$, i.e., the images of points of H' , under the morphism $B \xrightarrow{\phi_{t_\sigma^*M^2}} \mathbb{P}H^0(t_\sigma^*M^2) \simeq |t_\sigma^*M^2|, b \mapsto t_b^*\theta + t_{-b+2\sigma}^*\theta$ (due to Wirtinger), have their linear span as $|t_\sigma^*M^2|$. (Here we assume that M is symmetric and that θ is the unique symmetric divisor in $|M|$.)

To see this, we show

Proposition 1.3. *Let \mathcal{L} be an ample line bundle on a simple abelian variety Z of dimension g and consider the associated rational map $Z \xrightarrow{\phi_{\mathcal{L}}} \mathbb{P}H^0(\mathcal{L})$. Then any finite subgroup G of Z of order strictly greater than $h^0(\mathcal{L}) \cdot g!$, generates the linear system $\mathbb{P}H^0(\mathcal{L})$. More precisely, the points $\phi_{\mathcal{L}}(h)$ where h runs over all elements of G not in the base locus of \mathcal{L} span $\mathbb{P}H^0(\mathcal{L})$ (see Prop. 3.4).*

We then apply Proposition 1.3 to $\mathcal{L} = t_\sigma^*M^2$ to obtain bounds as asserted for a polarized abelian variety (A, L) in Theorem 1.2.

Notation. The varieties considered in this article are defined over the complex numbers.

- Let \mathcal{L} be an ample line bundle on an abelian variety Z of dimension g .
- 1. The *fixed group* of \mathcal{L} is the group $K(\mathcal{L}) = \{z \in Z : \mathcal{L} \simeq t_z^*\mathcal{L}\}, t_z : Z \longrightarrow Z, x \mapsto z + x$.
- 2. The *theta group* of \mathcal{L} is the group $\mathcal{G}(\mathcal{L}) = \{(z, \phi) : \mathcal{L} \xrightarrow{\phi} t_z^*\mathcal{L}\}$.
- 3. The *Weil form* $e^{\mathcal{L}} : K(\mathcal{L}) \times K(\mathcal{L}) \longrightarrow \mathbb{C}^*$ is the commutator map $(x, y) \mapsto x'y'x'^{-1}y'^{-1}$, for any lifts $x', y' \in \mathcal{G}(\mathcal{L})$ of $x, y \in K(\mathcal{L})$.
- 4. $h^0(\mathcal{L}) = \dim H^0(Z, \mathcal{L})$.
- 5. If G is a finite subgroup of Z , then $\text{Card}(G) = \text{order}(G)$.

2. SURJECTIVITY OF THE MAPS $\rho_r, r \geq 3$

Suppose \mathcal{L} is an ample line bundle on a g -dimensional abelian variety A . Consider the multiplication maps

$$H^0(\mathcal{L})^{\otimes r} \xrightarrow{\rho_r} H^0(\mathcal{L}^r), \text{ for } r \geq 2.$$

The main result of this section is the following.

Proposition 2.1. *Suppose \mathcal{L} is an ample line bundle on an abelian variety A . If the multiplication map ρ_2 is surjective, then ρ_r is surjective, for all $r \geq 3$.*

First, we recall

Proposition 2.2. *Suppose L and M are ample line bundles on an abelian variety A .*

1) The multiplication map

$$\sum_{\alpha \in U} H^0(L \otimes \alpha) \otimes H^0(M \otimes \alpha^{-1}) \longrightarrow H^0(L \otimes M)$$

is surjective, for any nonempty Zariski open subset U of $\text{Pic}^0(A)$.

2) If the multiplication map $H^0(L) \otimes H^0(M) \longrightarrow H^0(L \otimes M)$ is surjective, then the multiplication maps

$$(a) H^0(L) \otimes H^0(M \otimes \alpha) \longrightarrow H^0(L \otimes M \otimes \alpha)$$

and

$$(b) H^0(L \otimes \alpha^{-1}) \otimes H^0(M \otimes \alpha) \longrightarrow H^0(L \otimes M)$$

are also surjective, for α in some nonempty Zariski open subset U of $\text{Pic}^0(A)$.

Proof. 1) See [3], 7.3.3.

2) The proof is standard. □

Proof of Proposition 2.1. We prove by induction on r . Suppose the multiplication map $\rho_r : H^0(\mathcal{L})^{\otimes r} \longrightarrow H^0(\mathcal{L}^r)$ is surjective, for some $r \geq 2$.

Consider the composed multiplication map

$$H^0(\mathcal{L})^{\otimes r+1} \xrightarrow{\text{Id} \otimes \rho_r} H^0(\mathcal{L}) \otimes H^0(\mathcal{L}^r) \xrightarrow{\rho_{1,r}} H^0(\mathcal{L}^{r+1}).$$

To see the surjectivity of the map $\rho_{r+1} = \rho_{1,r} \circ (\text{Id} \otimes \rho_r)$ we need to show that the map $\rho_{1,r}$ is surjective.

Using Proposition 2.2 1), we can write

$$(*) \quad H^0(\mathcal{L}).H^0(\mathcal{L}^r) = \sum_{\alpha \in U} H^0(\mathcal{L}).H^0(\mathcal{L} \otimes \alpha^{-1}).H^0(\mathcal{L}^{r-1} \otimes \alpha)$$

for any nonempty Zariski open subset U of $\text{Pic}^0(A)$.

Since ρ_2 is surjective, by Proposition 2.2 2) (a), there exists a nonempty Zariski open subset U' of $\text{Pic}^0(A)$, such that for $\alpha^{-1} \in U'$,

$$(**) \quad H^0(\mathcal{L}).H^0(\mathcal{L} \otimes \alpha^{-1}) = H^0(\mathcal{L}^2 \otimes \alpha^{-1})$$

Now in (*), using (**) and again applying Proposition 2.2 1), we obtain

$$\begin{aligned} H^0(\mathcal{L}).H^0(\mathcal{L}^r) &= \sum_{\alpha^{-1} \in U'} H^0(\mathcal{L}^2 \otimes \alpha^{-1}).H^0(\mathcal{L}^{r-1} \otimes \alpha) \\ &= H^0(\mathcal{L}^{r+1}). \end{aligned}$$

□

3. SURJECTIVITY OF THE MAP ρ_2

Let Z be a g -dimensional abelian variety and let D be an ample divisor on Z . We denote $M = \mathcal{O}(D)$ to be the ample line bundle on Z . Let G be a finite subgroup of Z . Consider the homomorphism $\psi_M : Z \longrightarrow \text{Pic}^0(Z)$, $z \mapsto t_z^*(M) \otimes M^{-1}$. Let $G' \subset \text{Pic}^0(Z)$ be the image of G under this homomorphism. Consider a finite subgroup $J \subset \text{Pic}^0(Z)$ and containing the subgroup G' . Construct an étale cover $\pi : X \longrightarrow Z$ corresponding to J , which is of degree equal to $\text{Card} J$. Let $N = \mathcal{O}(\pi^{-1}D)$ be the ample line bundle on X .

Notice that if $h \in G \cap K(M)$, then $t_h^*M \simeq M$, and this implies that $D + h$ is linearly equivalent to D on Z . If $\psi_N : X \longrightarrow \text{Pic}^0(X)$ is the map $x \mapsto t_x^*N \otimes N^{-1}$ and $\hat{\pi} : \text{Pic}^0(Z) \longrightarrow \text{Pic}^0(X)$ is the dual of the map π , then since $\hat{\pi}(J) =$

$\{0\}$, $\pi^{-1}G \subset K(N) = \text{Ker}\psi_N$ (since $\psi_N = \hat{\pi} \circ \psi_M \circ \pi$). This means that the divisors $\pi^{-1}(D+h) \in |N|$, for all $h \in G$.

Choose the subgroup J such that N is base point free. (In fact, if J contains the subgroup of 3-torsion points of $\text{Pic}^0(Z)$ and G' , then, by the above discussion, $X_{[3]} \subset K(N)$, where $X_{[3]}$ is the subgroup of 3-torsion points of X . This implies, by [3] 2.5.6, that $N = K^3$, for some ample line bundle K on X and by a theorem of Lefschetz (see [3], 4.5.1), N is very ample.)

We will use the following.

Lemma 3.1. *Let V be a variety and $\mathcal{V} \subset \text{Div}(V)$ be an irreducible family of effective Cartier divisors D_t on V . Suppose $W = \bigcap_{t \in \mathcal{V}} D_t \subset V$ and is nonempty and $r = \text{codim}(W)$. Then there exist divisors D_j , $j = 1, 2, \dots, r$, in \mathcal{V} that intersect properly and $\dim W = \dim \bigcap_{i=1}^r D_i$.*

Proof. We use induction on j . Let D_1, D_2, \dots, D_j ($j < r$) be chosen in \mathcal{V} such that they intersect properly in V . Now write $D_1 \cap D_2 \cap \dots \cap D_j = G_1 \cup G_2 \cup \dots \cup G_s$, where G_1, \dots, G_s are irreducible components. Consider the closed subset $\mathcal{W}_i \subset \mathcal{V}$ parametrizing divisors that contain G_i for $i = 1, 2, \dots, s$. (Note that $\mathcal{W}_i \neq \mathcal{V}$, otherwise $G_i \subset W$, which is not possible since $\dim G_i > \dim W$.) Let U be the complement of $\bigcup_{i=1}^s \mathcal{W}_i$ in \mathcal{V} , which is nonempty since \mathcal{V} is irreducible. If $D_{j+1} \in U$, then $D_1 \cap \dots \cap D_j \cap D_{j+1}$ has codimension $j+1$ (communicated to us by A. Hirschowitz). \square

Remark 3.2. Suppose D_1, D_2, \dots, D_r are linearly equivalent effective divisors on a variety V , $W = \bigcap_{i=1}^r D_i$ and is nonempty and $r = \text{codim}(W)$. If \mathbb{P}^k denotes the span of the points D_i in the linear system $|D_1|$, then $W = \bigcap_{t \in \mathbb{P}^k} D_t$. Hence, by Lemma 3.1, there are r divisors $D_j \in \mathbb{P}^k$ that intersect properly.

With notation as above we have the following.

Proposition 3.3. *Let D be an ample divisor on a g -dimensional simple abelian variety Z . Let G be a finite subgroup of Z that is contained in D . Then $\text{Card}(G) \leq D^g$ (which equals $h^0(\mathcal{O}(D)) \cdot g!$, by the Riemann-Roch Theorem).*

Proof. We prove this in several steps.

Step 1: We reduce to the case when the divisors D and $D+h$, for all $h \in G$, are linearly equivalent and $\mathcal{O}(D)$ is base point free. Indeed, by the above discussion, choose a triple (X, N, π) , as above, corresponding to a subgroup $J \subset \text{Pic}^0(Z)$ such that N is base point free and $\psi_M(G) \subset J$. This shows that the divisors $\pi^{-1}D$ and $\pi^{-1}(D+h)$, for all $h \in G$, are linearly equivalent. Then we have a morphism $\phi_N : X \rightarrow \mathbb{P}H^0(N)$. Since π is a finite morphism of degree equal to $\text{Card}(J)$, by the projection formula, one sees that $\deg(\pi^{-1}W) = \text{Card}(J) \cdot \deg(W)$, for a subvariety W of Z . Since $(\pi^{-1}D)^g = \text{Card}(J) \cdot D^g$, if $\text{Card}(\pi^{-1}G) \leq (\pi^{-1}D)^g$, then $\text{Card}(G) \leq D^g$.

Step 2: We can now assume that D is an ample divisor on X and that $G \subset D$ is a finite subgroup such that D is linearly equivalent to $D+h$ for all $h \in G$ and $N = \mathcal{O}(D)$ is base point free. Let $Y = \bigcap_{h \in G} D+h$ and $s = \dim(Y)$. By Lemma 3.2, $Y \subset \bigcap_{j=1}^{g-s} D_j$ for some $g-s$ divisors $D_j \in |N|$ that intersect properly. Now $\deg(Y) = [Y] \cdot [D^s]$ (here $\deg(Y) = \deg(S)$, where $S \subset Y$ is of pure dimension s). Since $Y \subset \bigcap_{j=1}^{g-s} D_j$ we see that $\deg(Y) \leq D^g$. In particular, when $s = 0$, since $G \subset Y$, we get $\text{Card}(G) \leq D^g$.

Step 3: Suppose that $s > 0$. Let $Y = Y_1 \cup Y_2 \cup \dots \cup Y_r$, where Y_j , $1 \leq j \leq r$, are the irreducible components of Y such that $s = \dim Y_1 = \dim Y$. Then $\deg Y_1 \leq \deg Y$. Since Y is G -invariant, $\bigcup_{h \in G} Y_1 + h \subset Y$ and $\sum_{h \in \frac{G}{G_{Y_1}}} \deg(Y_1 + h) \leq \deg Y$, where $G_{Y_1} = \{h \in G : Y_1 + h = Y_1\}$ is a subgroup of G . Hence we get the inequalities $\text{Card}(\frac{G}{G_{Y_1}}) \cdot \deg Y_1 \leq \deg Y \leq D^g$, i.e., $\text{Card}(G) \leq \frac{\text{Card}(G_{Y_1})}{\deg Y_1} \cdot D^g$. To complete our proof, we need to show that $\text{Card}(G_{Y_1}) \leq \deg Y_1$.

Step 4: Now $G_{Y_1} \subset \text{Stab}(Y_1) = \{a \in X : Y_1 + a = Y_1\}$. Observe that $\text{Stab}(Y_1) = \bigcap_{y \in Y_1} Y_1 - y$. Now for a point $y_0 \in Y_1$, $\text{Stab}(Y_1) = (Y_1 - y_0) \bigcap_{y \in Y_1} Y_1 - y \subset (Y_1 - y_0) \bigcap_{h \in G, y \in Y_1} D + h - y$. Let $P = (Y_1 - y_0) \bigcap_{h \in G, y \in Y_1} D + h - y$. We proceed to show that $\deg(\text{Stab}(Y_1)) \leq \deg(P)$. This will be true if $\text{Stab}(Y_1)$ and P have the same dimension. Now, we have

$$\begin{aligned} \bigcap_{h \in G, y \in Y_1} D + h - y &= \bigcap_{y \in Y_1} Y - y \\ &= \bigcap_{y \in Y_1} ((Y_1 \cup Y_2 \cup \dots \cup Y_r) - y) \\ &= \left(\bigcap_{y \in Y_1} Y_1 - y \right) \cup \left(\bigcap_{y \in Y_1} Y_2 - y \right) \cup \dots \cup \left(\bigcap_{y \in Y_1} Y_r - y \right). \end{aligned}$$

(To see the above last equality: if $x \in \bigcap_{y \in Y_1} (Y_1 \cup Y_2 \cup \dots \cup Y_r) - y$, then $x + y \in Y_1 \cup Y_2 \cup \dots \cup Y_r$, $\forall y \in Y_1$. Via the translation map $Y_1 \rightarrow Y_1 \cup Y_2 \cup \dots \cup Y_r$, $y \mapsto y + x$ and since Y_1 is irreducible, $x + y \in Y_j$, for some j and for all $y \in Y_1$, i.e., $x \in \bigcap_{y \in Y_1} Y_j - y$ showing one way inclusion, the other inclusion being obvious.)

We now see that if $j \neq 1$ and $x \in \bigcap_{y \in Y_1} Y_j - y$, then $Y_1 + x \subset Y_j$. If $\dim Y_j < \dim Y_1$, then this is absurd and so $\bigcap_{y \in Y_1} Y_j - y$ is empty. If $\dim Y_j \geq \dim Y_1$, since Y_1 is of maximal dimension in Y , $\dim Y_j = \dim Y_1$ and $Y_1 + x = Y_j$. This implies that $\bigcap_{y \in Y_1} Y_j - y = \bigcap_{y \in Y_1} Y_1 + x - y = \text{Stab}(Y_1) + x$. Hence $\bigcap_{h \in G, y \in Y_1} D + h - y$, P and $\text{Stab}(Y_1)$ are of equal dimension, say equal to m and $\deg(\text{Stab}(Y_1)) \leq \deg P$.

Step 5: We proceed to show that $\deg(P) \leq \deg(Y_1)$. Consider the Poincaré line bundle \mathcal{P} on $X \times \text{Pic}^0(X)$. Let p_1 and p_2 denote the projections onto X and $\text{Pic}^0(X)$ respectively from $X \times \text{Pic}^0(X)$. Consider the sheaf $\mathcal{E} = p_{2*}(p_1^*N \otimes \mathcal{P})$ on $\text{Pic}^0(X)$. Since the vector spaces $H^0(N \otimes \alpha)$ are of constant dimension for all $\alpha \in \text{Pic}^0(X)$, by Grauert's theorem, \mathcal{E} is a vector bundle on $\text{Pic}^0(X)$. Let $\mathbb{P}(\mathcal{E})$ denote the associated projective bundle on $\text{Pic}^0(X)$. Consider the natural morphism $p_2^*(\mathcal{E}) \rightarrow p_1^*N \otimes \mathcal{P}$. This is surjective, since on any fibre $X \times \alpha$, $(p_1^*N \otimes \mathcal{P})_\alpha \simeq N \otimes \alpha$ which is globally generated (since N is globally generated) and $\mathcal{E}(\alpha) \simeq H^0(N \otimes \alpha)$. Hence this defines a morphism $\delta_N : X \times \text{Pic}^0(X) \rightarrow \mathbb{P}(\mathcal{E})$. Let $\mathbb{P}(\mathcal{E})^\vee$ denote the dual projective bundle over $\text{Pic}^0(X)$. In general, the parameter space $\mathcal{V} \subset \mathbb{P}(\mathcal{E})^\vee$ of the family $\{D + h - y\}_{h \in G, y \in Y_1}$ may not form an irreducible variety (unless $G_{Y_1} = G$), but we construct an irreducible subvariety $\mathcal{F} \subset \mathbb{P}(\mathcal{E})^\vee$ such that $\mathcal{V} \subset \mathcal{F}$ and $\bigcap_{h \in G, y \in Y_1} D + h - y = \bigcap_{t \in \mathcal{F}} D_t$, where D_t denotes the divisor corresponding to t in $\mathbb{P}(\mathcal{E})^\vee$ (**).

Step 6: Construction of \mathcal{F} :

Consider the subspace T of $H^0(X, N)$ spanned by sections s_h , $h \in G$ such that the divisor of s_h is $D + h$. Consider the addition map $m : X \times X \rightarrow X$, $(x, y) \mapsto x + y$. Recall the skew-Pontryagin product of the sheaves \mathcal{O}_X and N , $N \hat{*} \mathcal{O}_X = (p_1)_*(m^*N)$ (see [8], p. 653), where p_1 (resp. p_2) : $X \times X \rightarrow X$

denotes the first (resp. second) projection. Then, by Grauert's theorem, $N^*\mathcal{O}_X$ forms a vector bundle on X with fibres $(N^*\mathcal{O}_X)_x \simeq H^0(t_x^*N)$. By [8], Remark 1.2, $N^*\mathcal{O}_X \simeq N * \mathcal{O}_X$ where $N * \mathcal{O}_X = m_*(p_1^*N)$ is the Pontryagin product and by [5], p. 161, there are isomorphisms $\mathcal{O}_X \otimes H^0(X, N) \xrightarrow{f} N^*\mathcal{O}_X \simeq \psi_N^*\mathcal{E} \otimes N$ ($\psi_N : X \rightarrow \text{Pic}^0(X)$ is the isogeny $x \mapsto t_x^*N \otimes N^{-1}$). Consider the image F under f of the trivial subbundle $\mathcal{O}_X \otimes T$ in $N^*\mathcal{O}_X$. Then the fibre of F at $x \in X$ is the vector subspace of $H^0(t_x^*N)$ spanned by the sections $t_x^*s_h$ whose divisor is $D+h-x$, for $h \in G$. Now $\mathbb{P}(F)$ is a projective subbundle of $\mathbb{P}(\psi_N^*\mathcal{E} \otimes N) \simeq \mathbb{P}(\psi_N^*\mathcal{E})$ (since N is a line bundle). Since Y_1 is irreducible, the projective bundle $\mathbb{P}(F)$ restricted to Y_1 is an irreducible subvariety, and let \mathcal{F} be the image of this irreducible variety in $\mathbb{P}(\mathcal{E})$. Hence \mathcal{F} is irreducible and, by construction, if $R \in |F_y|$, $y \in Y_1$, then $\bigcap_{h \in G} D+h-y \subset R$ and \mathcal{F} satisfies property (**).

Step 7: By Lemma 3.1, there exist divisors $D_1, D_2, \dots, D_{g-m} \in \mathcal{F}$ such that $\bigcap_{h \in G, y \in Y_1} D+h-y \subset D_1 \cap D_2 \cap \dots \cap D_{g-m}$. Hence $P \subset (Y_1 - y_0) \cap D_1 \cap D_2 \cap \dots \cap D_{g-m} \subset D_1 \cap D_2 \cap \dots \cap D_{g-m}$. This implies that $\deg(P) \leq \deg(Y_1 - y_0)$, and by Step 2 and Step 4, $\deg \text{Stab}(Y_1) \leq \deg Y_1 \leq D^g$ (since by Step 2, $\deg(Y_1) \leq \deg(Y) \leq D^g$). Since X is simple, $\text{Stab}(Y_1)$ is zero-dimensional and $G_{Y_1} \subset \text{Stab}(Y_1)$ implies that $\text{Card}(G_{Y_1}) \leq \deg(Y_1)$. Hence by Step 3, $\text{Card}(G) \leq D^g$. This ends the proof. \square

This is equivalent to the following.

Proposition 3.4. *Let \mathcal{L} be an ample line bundle on a simple abelian variety Z and consider the associated rational map $Z \xrightarrow{\phi_{\mathcal{L}}} \mathbb{P}H^0(\mathcal{L})$. Then any finite subgroup G of Z , of order strictly greater than $h^0(\mathcal{L}) \cdot g!$, generates $\mathbb{P}H^0(\mathcal{L})$. More precisely, the points $\phi_{\mathcal{L}}(g)$ where g runs over all elements of G not in the base locus of \mathcal{L} span $\mathbb{P}H^0(\mathcal{L})$.*

We recall the following result, which we will need in the proof of Theorem 1.2.

Proposition 3.5 (Wirtinger). *Let (Z, Θ) be a principally polarized abelian variety and $\mathcal{L} = \mathcal{O}(\Theta)$ (here Θ is assumed to be a symmetric divisor). There is a nondegenerate inner product $R : H^0(\mathcal{L}^2) \otimes H^0(\mathcal{L}^2) \rightarrow \mathbb{C}$ (which is symmetric or skew-symmetric depending on whether the multiplicity of the zero element 0 on Θ , $\text{mult}_0\Theta$, is even or odd) such that if R induces the isomorphism R' ,*

$$\mathbb{P}(H^0(\mathcal{L}^2)) \simeq \mathbb{P}(H^0(\mathcal{L}^2)^*) = |2\Theta|,$$

then the composed morphism

$$Z \xrightarrow{\phi_{\mathcal{L}^2}} \mathbb{P}(H^0(\mathcal{L}^2)) \xrightarrow{R'} |2\Theta|$$

is the morphism

$$\phi : Z \rightarrow |2\Theta|, \quad x \mapsto \Theta_x + \Theta_{-x},$$

where Θ_x is the translate of Θ by x on Z .

Proof. See [6], Proposition, p. 335. \square

Proof of Theorem 1.2. Consider a polarized simple abelian variety (A, L) of dimension g such that $h^0(L) > 2^g \cdot g!$.

Consider the multiplication map

$$H^0(L) \otimes H^0(L) \xrightarrow{\rho_2} H^0(L^2).$$

This map factors via

$$\operatorname{Sym}^2 H^0(L) \xrightarrow{\rho_2} H^0(L^2).$$

Let $H \subset K(L)$ be a maximal isotropic subgroup for the Weil form e^L . Consider the isogeny $A \xrightarrow{\pi} B = \frac{A}{H}$. Then L descends down to a principal polarization M on B . We may assume that M is symmetric, i.e., $M \simeq i^*M$, $i(b) = -b, b \in B$. Using the fact that $\pi_*\mathcal{O}_A = \bigoplus_{\chi \in \hat{H}} L_\chi$, where L_χ denotes the degree 0 line bundle on B corresponding to the character χ on H , by the projection formula, $\pi_*L = \bigoplus_{\chi \in \hat{H}} M \otimes L_\chi$ and $\pi_*L^2 = \bigoplus_{\chi \in \hat{H}} M^2 \otimes L_\chi$. Hence we obtain the following decompositions:

$$H^0(L) = \bigoplus_{\chi \in \hat{H}} H^0(M \otimes L_\chi)H^0(L^2) = \bigoplus_{\chi \in \hat{H}} H^0(M^2 \otimes L_\chi).$$

Write $\operatorname{Sym}^2 H^0(L) = \sum_{\chi, \chi' \in \hat{H}} H^0(M \otimes L_{\chi'}) \cdot H^0(M \otimes L_{\chi \cdot \chi'^{-1}})$. Consider the multiplication maps

$$\sum_{\chi' \in \hat{H}} H^0(M \otimes L_{\chi'}) \cdot H^0(M \otimes L_{\chi \cdot \chi'^{-1}}) \xrightarrow{\rho_\chi} H^0(M^2 \otimes L_\chi).$$

Since $\rho_2 = \bigoplus_{\chi \in \hat{H}} \rho_\chi$, it will suffice to show the surjectivity of ρ_χ for each $\chi \in \hat{H}$.

Since the pair (B, M) is principally polarized, the homomorphism $\psi_M : B \longrightarrow \operatorname{Pic}^0(B)$ is an isomorphism. Let $H' = \psi_M^{-1}(\hat{H})$ and $\theta \in |M|$ be the unique symmetric divisor.

Case 1: Suppose χ is trivial.

We see that the surjectivity of the map ρ_{triv} is equivalent to showing that the reducible divisors $\theta_h + \theta_{-h}$ generate the linear system $|M^2|$, for $h \in H'$. By Proposition 3.5, using the morphism $\phi : B \longrightarrow |M^2|$, this is the same as saying that the image of the subgroup H' under the morphism ϕ_{M^2} generates the projective space $\mathbb{P}H^0(M^2)$.

Case 2: Suppose χ is nontrivial.

First, notice that if $b \in B$, then $\psi_{M^2}(b) = \psi_M(2b)$. Let $\sigma \in B$ be such that $\psi_{M^2}(\sigma) = L_\chi$, i.e., $\psi_M(2\sigma) = L_\chi$. Hence the map ρ_χ is surjective if the reducible divisors $\theta_h + \theta_{-h+2\sigma}$ span the linear system $|t_\sigma^*M^2|$ for $h \in H' = \psi_M^{-1}(\hat{H})$. Now if $b \in B$, then $\theta_b + \theta_{-b+2\sigma} = (\theta_\sigma)_{b-\sigma} + (\theta_\sigma)_{-b+\sigma}$, which is the image of the divisor $\theta_{b-\sigma} + \theta_{-b+\sigma}$ under the morphism $|M^2| \longrightarrow |t_\sigma^*M^2|$ given by the translation map $B \xrightarrow{t_\sigma} B$. Hence the morphism $\phi_\sigma : A \longrightarrow |t_\sigma^*M^2|$ is given as $b \mapsto \theta_b + \theta_{-b+2\sigma}$. This implies that ρ_χ is surjective if and only if the points in $\phi_\sigma(H')$ generate the linear system $|t_\sigma^*M^2|$.

Since the pair (A, L) is a simple polarized abelian variety with $h^0(L) = \operatorname{Card}(H') > 2^g \cdot g! = h^0(t_\sigma^*M^2) \cdot g!$, by Proposition 3.4, ρ_χ is surjective for all $\chi \in \hat{H}$. Hence, by Proposition 2.1, our proof is now complete. \square

Remark 3.6. 1) Suppose $g = 1$. Then any line bundle of degree strictly greater than 2 on an elliptic curve gives a projectively normal embedding. Hence the bound is sharp.

2) Suppose $g = 2$. If $L \simeq N^2$, where N is an ample symmetric line bundle with $h^0(N) = 2$ on an abelian surface A , then it follows that $h^0(L) = 8$ (in terms of “type” of an ample line bundle, N is of type (1,2) and hence L is of type (2,4) and $h^0(L) = 8$). By [3], 10.1.4, N has 4 base points, say x_1, x_2, x_3 and x_4 , which are 4-torsion points on A and, moreover, $2x_i \in K(N) = \operatorname{Ker} \psi_N$ where

$\psi_N : A \longrightarrow \text{Pic}^0(A)$, $a \mapsto t_a^* N \otimes N^{-1}$. Let $\alpha_i = \psi_N(x_i)$, for $i = 1, 2, 3, 4$. Now the points x_i are base points for N , for $i = 1, 2, 3, 4$, is equivalent to saying that the origin $0 \in A$ is a base point for $N \otimes \alpha_i$, for $i = 1, 2, 3, 4$. Also $2x_i \in K(N)$ implies that the points α_i are 2-torsion points in $\text{Pic}^0(A)$. Hence by Ohbuchi's Theorem 1.1, L does not give a projectively normal embedding. So the bound is sharp.

3) Suppose $g = 3$. If $L \simeq N^3$, where N is a principal polarization on an abelian threefold A , then $h^0(L) = 27$. But by Koizumi's Theorem, L gives a projectively normal embedding. So the bound is not sharp in this case.

ACKNOWLEDGEMENTS

We thank M. Hindry and K. Paranjape for useful conversations and A. Beauville for pointing out a gap in Prop. 3.3 in an earlier version. We are grateful to A. Hirschowitz for his helpful suggestions incorporated here. This work was done at the University of Paris-6 and the University of Essen. Their hospitality and support from the French Ministry of Education, Research and Technology and DFG "Arithmetik und Geometrie" Essen, is gratefully acknowledged.

REFERENCES

- [1] Iyer, J.: *Projective normality of abelian surfaces given by primitive line bundles*, Manuscripta Math., **98**, 139-153 (1999). MR **2000b**:14056
- [2] Koizumi, S.: *Theta relations and projective normality of abelian varieties*, American Journal of Mathematics, **98**, 865-889 (1976). MR **58**:702
- [3] Lange, H. and Birkenhake, Ch.: *Complex abelian varieties*, Grundlehren der Mathematischen Wissenschaften, **302**, Springer-Verlag, Berlin, (1992). MR **94j**:14001
- [4] Lazarsfeld, R.: *Projectivité normale des surfaces abéliennes*, Rédigé par O. Debarre. Prépublication No. **14**, Europroj- C.I.M.P.A., Nice, (1990).
- [5] Mukai, S.: *Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves*, Nagoya Math. J., **81**, 153-175 (1981). MR **82f**:14036
- [6] Mumford, D.: *Prym varieties I*, in: Contributions to Analysis (a collection of papers dedicated to Lipman Bers), Academic Press, New York, 325-350 (1974). MR **52**:415
- [7] Ohbuchi, A.: *A note on the normal generation of ample line bundles on abelian varieties*, Proc. Japan Acad. Ser. A Math. Sci. **64**, 119-120 (1988). MR **90a**:14062a
- [8] Pareschi, G.: *Syzygies of abelian varieties*, J. Amer. Math. Soc. **13**, 651-664 (2000). MR **2001f**:14086

MAX-PLANCK-INSTITUT FÜR MATHEMATIK, VIVATSGASSE 7, D-53111, BONN, GERMANY
 E-mail address: jniyer@mpim-bonn.mpg.de

CLUSTERING OF CRITICAL POINTS IN LEFSCHETZ FIBRATIONS AND THE SYMPLECTIC SZPIRO INEQUALITY

V. BRAUNGARDT AND D. KOTSCHICK

ABSTRACT. We prove upper bounds for the number of critical points in semi-stable symplectic Lefschetz fibrations. We also obtain a new lower bound for the number of nonseparating vanishing cycles in Lefschetz pencils and reprove the known lower bounds for the commutator lengths of Dehn twists.

1. INTRODUCTION

It is a result of Szpiro [16] that in a semistable algebraic family of elliptic curves over \mathbb{CP}^1 the number of critical points is bounded above by six times the number of singular fibers. In fact, Szpiro considered the arithmetic situation over a number field, and the function field case was just a byproduct of these considerations. Recently, Beauville [3] proved a generalisation of Szpiro's inequality to fibered surfaces, where both the base and the fiber have arbitrary genus.

Amorós et al. [1] gave a group-theoretic proof of Szpiro's inequality for semistable symplectic Lefschetz pencils, and this was extended to certain hyperelliptic semistable symplectic Lefschetz pencils by Bogomolov et al. [4]. Their result is that if all the vanishing cycles are non-separating and the fibration has a topological section, then the number of critical points is bounded above by $4h + 2$ times the number of singular fibers, where h is the genus of a smooth fiber.

The purpose of this paper is to prove an inequality of Szpiro type for all semistable symplectic Lefschetz fibrations of arbitrary base and fiber genus, without any assumption on the vanishing cycles and without the hyperelliptic assumption. In the case of pencils our result is that the number of critical points is bounded above by $6h(D - 1)$, where h is the genus of a smooth fiber and D is the number of singular fibers. See Theorem 15 and Remark 18 below. In genus one we once more recover the complex function field version of Szpiro's theorem. In higher genus, our inequality is weaker than the one obtained by Beauville [3] in the algebraic situation.

The point of the symplectic Szpiro inequality is that while one can always perturb a Lefschetz fibration in the neighbourhood of a singular fiber so as to make it

Received by the editors September 10, 2002.

2000 *Mathematics Subject Classification*. Primary 57R17, 57R57, 14H10.

Support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged. The authors are members of the *European Differential Geometry Endeavour* (EDGE), Research Training Network HPRN-CT-2000-00101, supported by The European Human Potential Programme.

injective on its critical set, there are global obstructions to the clustering or concentration of critical points in a fiber. These obstructions are essentially the ones discovered by Endo and Kotschick [6] in their proof that the commutator lengths of powers of Dehn twists have linear growth.

In Section 3 we generalise the main result of [6] to relatively minimal Lefschetz fibrations which need not be injective on their critical sets, and then deduce a Szpiro inequality for fibrations over bases of positive genus from this generalisation. We then rederive the lower bounds for the commutator lengths of Dehn twists proved in [6], [8]. There the argument with the Kneser inequality from [9] was combined with the observation that iteration of Dehn twists makes the signatures of Lefschetz fibrations more and more negative. This is clear for separating Dehn twists, but also works for nonseparating Dehn twists by using a handle decomposition, cf. [8]. Here we give a treatment that avoids handle decompositions and derives the growth of the negative definite part of the intersection form from the purely homological Proposition 5 which is standard in algebraic geometry but holds also in our symplectic setup. This makes the argument self-contained; in particular, it is independent of signature calculations for Lefschetz fibrations carried out ad hoc or by using the Meyer signature cocycle.

In Section 4 we generalise results of Li [10] and Stipsicz [14], [15] to relatively minimal symplectic Lefschetz fibrations which need not be injective on their critical sets, and then derive a Szpiro inequality from these generalisations. We also prove that the number n of nonseparating vanishing cycles in a Lefschetz pencil of genus h is no less than $\frac{1}{5}(8h - 3)$. The best previously known bound was $n \geq h$, cf. Remark 22. Here too our argument is independent of any signature calculations.

2. SEMISTABLE SYMPLECTIC LEFSCHETZ FIBRATIONS

For definitions and background on differentiable Lefschetz fibrations we refer to [7].

Let $f: X \rightarrow B$ be an oriented Lefschetz fibration with base genus g , fiber genus $h \geq 1$, with n nonseparating and s separating vanishing cycles. We denote by $k = n + s$ the total number of vanishing cycles and by D the number of singular fibers, so that $k \geq D$. We denote by N the total number of components of singular fibers. Note that for $h = 1$ we have $k = N$.

We shall assume throughout that the total space X is a symplectic manifold in such a way that the fibers are symplectic submanifolds. By a theorem of Gompf, cf. [7], this is no restriction if $h \geq 2$.

The Euler characteristic of a Lefschetz fibration is given by

(1)
$$\chi(X) = 4(g - 1)(h - 1) + k .$$

The following is elementary:

Lemma 1. *If K denotes the canonical class of an almost complex structure associated with the symplectic structure, then*

(2)
$$\begin{aligned} K^2 &= 5\chi(X) - 6 + 6b_1(X) - 6b_2^-(X) \\ &= 20(g - 1)(h - 1) + 5k - 6 + 6b_1(X) - 6b_2^-(X) . \end{aligned}$$

Definition 2. A symplectic Lefschetz fibration $f: X \rightarrow B$ is called *semistable* if every 2-sphere component of a singular fiber contains at least two critical points.

The name comes from the fact that if we choose a Riemannian metric on X and consider the induced conformal structures on the fibers, these become semistable algebraic curves if and only if the above topological condition is satisfied.

Lemma 3. *A symplectic Lefschetz fibration is semistable if and only if it is relatively minimal.*

Proof. Assume it is semistable and suppose that a singular fiber has more than one component. Let S be an irreducible component. If F denotes the class of the generic fiber, then $F \cdot S = 0$. Thus,

$$S^2 = S \cdot (S - F) \leq -1$$

since the singular fiber is connected, and different components intersect positively unless they are disjoint. If $S \cdot (S - F) = -1$, then by the definition of semistability S is not an embedded sphere. Thus there is no embedded sphere of self-intersection -1 in any fiber.

Conversely, if the fibration is relatively minimal, then by the same calculation as above, there can be no spherical component containing only one critical point. \square

A variation of this calculation shows the following:

Lemma 4. *Let X be a Lefschetz fibration and F_0 a singular fiber. Then $H_2(X)$ contains a negative definite subspace for the intersection form spanned freely by the classes of all the components of F_0 but one.*

Proof. In the case of algebraic surfaces, this is well known as Zariski's lemma; see e. g. [2]. The argument given there goes through in the symplectic situation because the components of singular fibers that we have in the statement intersect each other positively if they are not disjoint. \square

Proposition 5. *Let X be any Lefschetz fibration. Then*

$$(3) \quad b_2^-(X) \geq 1 + N - D .$$

Proof. The negative definite subspaces of $H_2(X)$ obtained by applying the preceding lemma to the different singular fibers are mutually orthogonal. Thus the direct sum of all these subspaces is negative definite of dimension $N - D$, and is still orthogonal to the class of a generic fiber, which has zero self-intersection. \square

We shall also need the following estimates for the number of components of singular fibers:

Proposition 6. *For every Lefschetz fibration we have*

$$(4) \quad N \geq s + D ,$$

$$(5) \quad N \geq k - (h - 1)D .$$

Proof. The first inequality is immediate from the definition.

To prove (5), let Σ be the union of the singular fibers. We can calculate the Euler characteristic of Σ by comparing the singular fibers to a generic fiber F ,

$$\chi(\Sigma) = D\chi(F) + k = -2(h - 1)D + k ,$$

or by summing over all components C_1, \dots, C_N of the singular fibres,

$$\chi(\Sigma) = \sum_{i=1}^N \chi(C_i) - k \leq 2N - k.$$

Thus we have $N \geq k - (h - 1)D$. \square

Remark 7. For fibrations with only stable fibers we have $k \leq 3(h - 1)D$ and $N \leq 2(h - 1)D$ for purely topological reasons.

3. LEFSCHETZ FIBRATIONS OVER BASES OF POSITIVE GENUS

In this section we prove the main technical result about relatively minimal Lefschetz fibrations and derive inequalities of Szpiro type under the assumption that the base has positive genus.

Theorem 8. *Let X be a connected smooth closed oriented 4-manifold and $f: X \rightarrow B$ a relatively minimal symplectic Lefschetz fibration with fiber genus $h \geq 1$ and base genus $g \geq 1$ having k vanishing cycles, D singular fibers and N irreducible components of singular fibers. Then*

$$(6) \quad 5k + 6(3h - 1)(g - 1) \geq 6(N - D).$$

Proof. Since X is assumed to be relatively minimal, the positivity of the base genus implies that X is minimal and not ruled, because any pseudo-holomorphic sphere in X would have to be contained in a fiber. Thus Liu's extension [11] of Taubes's results [17] implies $K^2 \geq 0$, which we can write as

$$(7) \quad b_2^+(X) \geq \frac{1}{5}(b_2^-(X) + 4b_1(X) - 4).$$

Using (3) and $b_1(X) \geq 2g \geq 2$, we obtain

$$b_2^+(X) \geq 1 + \frac{1}{5}(N - D).$$

Since the claim (6) is trivial for $N = D$, we may assume $N - D \geq 1$, and therefore $b_2^+(X) \geq 2$.

Since X is minimal with $b_2^+(X) \geq 2$, we can use the result of Taubes [17] to obtain a symplectically embedded surface $\Sigma \subset X$ representing the canonical class K of X . It may be disconnected, but because X is minimal, Σ has no spherical component. In the argument below we will tacitly assume that it is connected. In the general case the same argument works by summing over the components.

The genus of Σ is given by the adjunction formula $g(\Sigma) - 1 = K^2$. The fibration f induces a smooth map $\Sigma \rightarrow B$ of degree d equal to the algebraic intersection number of Σ with a fiber. This is calculated from the adjunction formula applied to a smooth fiber F , which is a symplectic submanifold:

$$(8) \quad d = \Sigma \cdot F = K \cdot F = 2h - 2.$$

Thus Kneser's inequality $g(\Sigma) - 1 \geq |d|(g(B) - 1)$ gives

$$K^2 \geq 2(h - 1)(g - 1).$$

Combining this with (2) and estimating K^2 from above using $b_2^-(X) \geq 1 + N - D$ and $b_1(X) \leq 2g + 2h$ we obtain

$$6(N - D) \leq 18(h - 1)(g - 1) - 12 + 12g + 12h + 5k.$$

Pulling the fibration back to large degree covers of the base B and applying the above inequality, we finally obtain (6). \square

Corollary 9. *Let X be a connected smooth closed oriented 4-manifold and $f: X \rightarrow B$ a relatively minimal symplectic Lefschetz fibration with fiber genus $h \geq 1$ and base genus $g \geq 1$ having s separating and n nonseparating vanishing cycles, D singular fibers and N irreducible components of singular fibers. Then the following inequalities hold:*

$$(9) \quad s \leq 6(3h - 1)(g - 1) + 5n ,$$

$$(10) \quad k \leq 6(3h - 1)(g - 1) + 6hD ,$$

$$(11) \quad N \leq 6(3h - 1)(g - 1) + (5h + 1)D .$$

Proof. The first claim follows from (6) using $N \geq s + D$. The second and third claim follow similarly using $k \leq N + (h - 1)D$. \square

Remark 10. The inequality (9) was originally proved by Endo and Kotschick [6] under the assumption $k = D$.

Remark 11. Note that for $h = 1$ we have $s = 0$ and $k = N$. From (10) or (11) we obtain

$$k = N \leq 12(g - 1) + 6D ,$$

which is a generalisation of the Szpiro inequality to semistable symplectic Lefschetz fibrations over bases of positive genus. For $h \geq 2$, either (10) or (11) can be regarded as a Szpiro-type inequality.

We now apply the previous discussion to give a new proof for the known lower bounds for the commutator lengths of powers of Dehn twists in mapping class groups.

Theorem 12. *Let a be a homotopically nontrivial simple closed curve on a surface F of genus $h \geq 2$, and let t_a be the corresponding Dehn twist. Suppose that t_a^k with $k > 0$ can be written as a product of g commutators. Then*

$$(12) \quad g \geq 1 + \frac{k}{6(3h - 1)} .$$

Proof. Consider the standard holomorphic Lefschetz fibration over the 2-disk D^2 with precisely one singular fiber F_0 with vanishing cycle a . Pulling back under the base change $z \mapsto z^k$ and taking the minimal resolution, we obtain a holomorphic Lefschetz fibration over D^2 with only one singular fiber having k vanishing cycles that are parallel copies of a . The monodromy of this fibration around the boundary of the disk is t_a^k . If this can be expressed as a product of g commutators, then we can find a smooth surface bundle with fiber F over a surface of genus g with one boundary component and the same restriction to the boundary. Let X be the Lefschetz fibration over the closed surface B of genus g obtained by gluing together the two fibrations along their common boundary.

By construction, X is symplectic and relatively minimal, so that we can apply Theorem 8. We have $D = 1$ and $N = k + 1$ if a is separating and $N = k$ if a is nonseparating. Theorem 8 gives $k \leq 6(3h - 1)(g - 1) + c$ with $c = 0$ or $c = 6$ depending on whether a is separating or not. In the latter case, by pulling back the

fibration to large degree coverings of the base, we also obtain $k \leq 6(3h - 1)(g - 1)$ as claimed. \square

Remark 13. The inequality (12) was originally proved by Endo and Kotschick [6] under the assumption that a is separating. In the context of Lefschetz fibrations with $k = D$ they used the separating assumption to conclude $b_2^-(X) \geq k + 1$ whenever a occurs k times as a vanishing cycle, since every separating vanishing cycle makes a negative contribution to the signature. Korkmaz [8] then observed that using a handle decomposition one still gets $b_2^-(X) \geq k$ in the nonseparating case, so that the argument goes through. Phrased as above, the proof works directly for both cases, since the required lower bound for $b_2^-(X)$ arises from the homological argument in Proposition 5.

4. LEFSCHETZ PENCILS

In this section we consider the case of pencils, i.e., Lefschetz fibrations over the 2-sphere. We shall assume that our pencils are nontrivial, meaning that they have at least one critical point each.

As in [10], [15], the case of ruled surfaces has to be considered separately.

Proposition 14. *Suppose X is the blowup in b points of a 2-sphere bundle over a surface of genus a , and that X admits a nontrivial relatively minimal symplectic Lefschetz pencil with fiber genus $h \geq 1$ having s separating and n nonseparating vanishing cycles, D singular fibers and N irreducible components of singular fibers. Then the following inequalities hold:*

$$(13) \quad k \geq 2h - 2 + \frac{3}{2}(N - D) ,$$

$$(14) \quad n \geq 2h - 2 + \frac{1}{2}(N - D) .$$

Proof. By the assumptions on X , we have $b_2^+ = 1$, $b_2^- = 1 + b$ and $b_1 = 2a$.

Stipsicz [14] proved that for a nontrivial relatively minimal Lefschetz pencil, $K^2 \geq 4(1 - h)$. His proof was written under the assumption that Lefschetz pencils are injective on their critical sets, but this can be achieved by perturbation, and the inequality involves only topological invariants which do not change under perturbation (unlike D and N); so the inequality is true in our case. Substituting the above numbers into it, we obtain

$$(15) \quad 4a \leq 2 + 2h - \frac{1}{2}b .$$

Computing the Euler characteristic of X in two different ways, we see that $4(1 - a) + b = \chi(X) = 4(1 - h) + k$, and so

$$k = b + 4h - 4a \geq 2h - 2 + \frac{3}{2}b = 2h - 2 + \frac{3}{2}(b_2^- - 1) \geq 2h - 2 + \frac{3}{2}(N - D) ,$$

where we have used first (15) and then Proposition 5. Thus we have proved the first claim. The second one follows from the first using Proposition 6. \square

In general we have:

Theorem 15. *Let X be a connected smooth closed oriented 4-manifold and $f: X \rightarrow S^2$ a nontrivial relatively minimal symplectic Lefschetz pencil with fiber genus $h \geq 1$*

having s separating and n nonseparating vanishing cycles, D singular fibers and N irreducible components of singular fibers. Then the following inequalities hold:

$$(16) \quad 5k \geq 6h + 6(N - D) ,$$

$$(17) \quad 5n \geq 6h + s ,$$

$$(18) \quad k \leq 6h(D - 1) ,$$

$$(19) \quad N \leq (5h + 1)(D - 1) - (h - 1) .$$

Proof. First we fiber sum X with a genus h bundle over the 2-torus and apply (9) to the resulting Lefschetz fibration. This shows that, since X is nontrivial, we must have $n > 0$. Therefore $b_1(X) \leq 2h - 1$.

Assuming first that X is not rational or ruled, we would like to use Taubes's result [17] as extended by Liu [11] to obtain $K^2 \geq 0$. The problem with this argument is that in the case of base genus zero, relative minimality does not imply minimality, so that Taubes's result is not available. However, if X is not rational or ruled, we have Li's inequality [10]

$$(20) \quad K^2 \geq 2 - 2h .$$

Li's argument assumes that the Lefschetz pencil is injective on its critical set, but this can be achieved by perturbation without affecting the inequality.

Using this, we proceed as in the proof of Theorem 8. Combining (20) with (2) we obtain (16) by using $g = 0$, $b_1(X) \leq 2h - 1$ and $b_2^-(X) \geq 1 + N - D$.

It remains to deal with the case that X is a (blowup of a) ruled surface. If $h \geq 3$, then (16) follows from (13).

If $h = 2$, suppose we have a fibration satisfying (13) but failing (16). Then we conclude that $k = 2$, giving $\chi(X) = 4(1 - h) + k = -2$. On the other hand, in the notation of Proposition 14 we have $a \leq \frac{1}{2}h + \frac{1}{2} - \frac{1}{8}b$, which gives $a \leq 1$. We conclude that $\chi(X) = 4(1 - a) + b \geq 0$, which is a contradiction.

If $h = 1$ for a fibration X satisfying (13) but failing (16), then we conclude that $k \leq 3$ and therefore $\chi(X) = k \leq 3$. As above we also obtain $a \leq 1$. If $a = 0$, then $\chi(X) = 4$, which is a contradiction. If $a = 1$, then (15) gives $b = 0$ and therefore $k = \chi(X) = 0$. But then the fibration is trivial.

Thus (16) is proved in all cases. Once we have (16), the other inequalities follow as in the proof of Corollary 9. \square

Remark 16. For non-ruled total spaces, the inequality (17) was proved by Li [10] under the assumption $k = D$.

Remark 17. If $k \neq 0$, we conclude that $D \geq 2$ from (18). Thus, the critical points of a nontrivial Lefschetz pencil can never be concentrated in a single fiber, compare [12].

Remark 18. Note that for $h = 1$ we have $k = N$ and both (18) and (19) reprove Szpiro's inequality [16] and give extensions to pencils with $h > 1$. Concerning (18), note that Bogomolov et al. [4] proved that $k \leq (4h + 2)D$ for hyperelliptic fibrations under the additional assumption that all vanishing cycles are nonseparating and that the fibration admits a topological section.

Remark 19. In the proof of Theorem 15 we have used $b_1(X) \leq 2h - 1$. If we actually know the first Betti number, then we obtain better inequalities.

The following theorem gives new bounds on the number of nonseparating vanishing cycles in Lefschetz pencils.

Theorem 20. *Let X be a connected smooth closed oriented 4-manifold and $f: X \rightarrow S^2$ a nontrivial relatively minimal symplectic Lefschetz pencil with fiber genus $h \geq 1$ having s separating and n nonseparating vanishing cycles, D singular fibers and N irreducible components of singular fibers. If X is not rational or ruled, then the following inequalities hold:*

$$(21) \quad 5k \geq 8h - 3 + 5(N - D) ,$$

$$(22) \quad 5n \geq 8h - 3 .$$

Proof. The assumption that f is relatively minimal again does not imply that X is minimal, in which case we would have $K^2 \geq 0$. However, $K^2 \geq 0$ can be rewritten as

$$(23) \quad 5b_2^+ - 4b_1 + 4 \geq b_2^- .$$

Blowing up or down does not change the left-hand side of this inequality; so, regardless of whether X is minimal or not, the fact that it is not rational or ruled implies

$$(24) \quad 5b_2^+(X) - 4b_1(X) + 4 \geq 0$$

because of [17] and [11]. Using (1) and

$$b_2^+(X) = b_2(X) - b_2^-(X) = \chi(X) - 2 + 2b_1(X) - b_2^-(X) ,$$

we obtain

$$(25) \quad 5k \geq 20h - 14 - 6b_1(X) + 5b_2^- .$$

As in the proof of Theorem 15 we have $b_1(X) \leq 2h - 1$. From Proposition 5 we have $b_2^-(X) \geq N - D + 1$. Thus (25) gives (21).

Using $N - D \geq s$, we obtain (22) from (21). □

Finally, we extend (22) to all Lefschetz pencils:

Theorem 21. *If $f: X \rightarrow S^2$ is a nontrivial Lefschetz pencil of genus $h \geq 1$, then f has at least $\frac{1}{5}(8h - 3)$ nonseparating vanishing cycles.*

Proof. Clearly it suffices to prove the relatively minimal case. If X is not rational or ruled, then the result is part of the previous Theorem, cf. (22). If X is rational or ruled, we have (14), which is enough as long as $h \geq 4$. On the other hand, if $h \leq 2$, then (17) implies the claim. Thus it remains to deal with the case $h = 3$ for ruled manifolds.

If $h = 3$ and X is the blowup of a ruled surface satisfying (14) but failing $n \geq \frac{1}{5}(8h - 3)$, then again $s \leq N - D = 0$, and $k = n = 4$. The same arguments as in the proof of Theorem 15 above then give $b = 0$ and $a = 2$. Thus we have a ruled surface without any blowups over a genus 2 surface, in which the generic fiber F of the Lefschetz fibration represents a homology class of zero self-intersection. Since F is a symplectic submanifold, the proof of the Thom conjecture implies that F has minimal genus in its homology class. We now show that this leads to a contradiction.

If X is the product ruled surface $S^2 \times \Sigma_2$, then the homology classes of zero self-intersection are the multiples of the two factors. The multiples of the first factor

are all represented by spheres, so F could only be a multiple of the second factor. Since it has larger genus than the second factor, it would represent d times the second factor with $|d| \geq 2$. But the projection to the second factor induces a map of degree d from F and from the normalisation of a singular fiber to Σ_2 . Since the normalisation of a singular fiber has genus ≤ 2 , it does not map to Σ_2 with degree ≥ 2 .

For the nontrivial ruled surface, F can also not be a multiple of the class of the 2-sphere in the ruling, and must therefore be a class that has intersection number $d \neq 0, \pm 1$ with the class of the sphere in the ruling. Thus it maps to the base surface of genus 2 with degree ≥ 2 , and we obtain the same contradiction as in the product case. \square

Remark 22. Theorem 6.1 of Stipsicz's paper [15] gives the lower bound $\frac{1}{5}(8h - 4)$ for the total number of critical points in nontrivial Lefschetz pencils (which are injective on their critical sets). However, the proof given there does not seem to cover the case of ruled surfaces with fibrations of small fiber genus, because it appeals to Theorem 1.1 which gives nothing for fiber genus ≤ 5 , for example. More importantly, Stipsicz's argument appeals to Theorem 1.4 of that paper, whose proof is incomplete, because it uses a statement about the fundamental group of a Lefschetz fibration which is not known to be true. In fact, it is false if one considers achiral Lefschetz fibrations; compare [1], p. 503.

In any case, the important difference between the result of Stipsicz and ours is the fact that we obtain a bound growing with $\frac{8}{5}h$ for the number of nonseparating vanishing cycles, rather than the total number of separating and nonseparating ones. In the nonseparating case the best previous result is that of Li [10], which is $n \geq h$.

Remark 23. Combining Theorems 15 and 21, the number of nonseparating vanishing cycles in a Lefschetz pencil is bounded below by

$$(26) \quad 5n \geq ts + (8 - 2t)h + 3(t - 1)$$

for all $t \in [0, 1]$.

5. FINAL COMMENTS

As in [9], [6], the arguments of this paper rely on the work of Taubes [17] in Seiberg-Witten theory, showing that for a minimal symplectic 4-manifold with $b_2^+ > 1$, the canonical class is represented by a symplectically embedded surface without spherical components. Recently, the methods pioneered in Donaldson's work on Lefschetz pencils have been applied successfully to reprove Taubes's result. Donaldson and Smith [5] did this under the additional assumption $b_2^+ - b_1 > 1$. This is not sufficient for our purposes, but, as explained by Smith at the end of [13], the arguments of [5] can be pushed to cover all cases where $b_2^+ > 2$.

This means that as long as we work with manifolds with $b_2^+ > 2$, our results can be proved independently of gauge theory. In Section 3 we appealed to Liu's work [11] in gauge theory for the case $b_2^+ = 1$, but this can be avoided. Therefore, our results on fibrations over bases of positive genus do not require any input from gauge theory. In the case where the base genus is zero, it is possible that $b_2^+ = 1$ or 2, but if we exclude those cases, then the proofs can be based on [5], [13] instead of [17], [11].

ACKNOWLEDGEMENT

We are grateful to L. Katzarkov for getting us interested in symplectic analogs of the Szpiro inequality, and for useful conversations. We also like to thank C. Bohr for useful discussions, and A. Stipsicz for reference [15].

REFERENCES

1. J. Amorós, F. Bogomolov, L. Katzarkov and T. Pantev, *Symplectic Lefschetz fibrations with arbitrary fundamental groups*, J. Differential Geometry **54** (2000), 489–545. MR **2002g**:57051
2. A. Beauville, *Complex algebraic surfaces*, Cambridge University Press, 1983. MR **85a**:14024
3. A. Beauville, *The Szpiro inequality for higher genus fibrations*, Preprint arXiv:math.AG/0109080.
4. F. Bogomolov, L. Katzarkov and T. Pantev, *Hyperelliptic Szpiro inequality*, Preprint arXiv:math.GT/0106212.
5. S. K. Donaldson and I. Smith, *Lefschetz pencils and the canonical class for symplectic 4-manifolds*, Preprint arXiv:math.SG/0012067.
6. H. Endo and D. Kotschick, *Bounded cohomology and non-uniform perfection of mapping class groups*, Invent. Math. **144** (2001), 169–175. MR **2001m**:57046
7. R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Math., vol. 20, Amer. Math. Soc., Providence, RI, 1999. MR **2000h**:57038
8. M. Korkmaz, *Commutators in mapping class groups and bounded cohomology*, Preprint arXiv:math.GT/0012162.
9. D. Kotschick, *Signatures, monopoles and mapping class groups*, Math. Research Letters **5** (1998), 227–234. MR **99d**:57023
10. T.-J. Li, *Symplectic Parshin-Arakelov inequality*, Internat. Math. Research Notices **2000**, 941–954. MR **2001i**:57038
11. A.-K. Liu, *Some new applications of the general wall crossing formula, Gompf's conjecture and its applications*, Math. Research Letters **3** (1996), 569–585. MR **97k**:57038
12. I. Smith, *Geometric monodromy and the hyperbolic disc*, Quart. J. Math. **52** (2001), 217–228. MR **2002c**:57046
13. I. Smith, *Serre-Taubes duality for pseudoholomorphic curves*, Preprint arXiv:math.SG/0106220.
14. A. I. Stipsicz, *On the number of vanishing cycles in Lefschetz fibrations*, Math. Research Letters **6** (1999), 449–456. MR **2000g**:57046
15. A. I. Stipsicz, *Singular fibers in Lefschetz fibrations on manifolds with $b_2^+ = 1$* , Topology Appl. **117** (2002), 9–21. MR **2002j**:57048
16. L. Szpiro, *Discriminant et conducteur des courbes elliptiques*, Astérisque **183** (1990), 7–18. MR **91g**:11059
17. C. H. Taubes, *$SW \Rightarrow Gr$: from the Seiberg-Witten equations to pseudo-holomorphic curves*, Jour. Amer. Math. Soc. **9** (1996), 845–918. MR **97a**:57033

MATHEMATISCHES INSTITUT, LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN, THERESIENSTRASSE 39, 80333 MÜNCHEN, GERMANY

E-mail address: Volker.Braungardt@mathematik.uni-muenchen.de

MATHEMATISCHES INSTITUT, LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN, THERESIENSTRASSE 39, 80333 MÜNCHEN, GERMANY

E-mail address: dieter@member.ams.org

ON MEASURES OF MAXIMAL AND FULL DIMENSION FOR POLYNOMIAL AUTOMORPHISMS OF \mathbb{C}^2

CHRISTIAN WOLF

ABSTRACT. For a hyperbolic polynomial automorphism of \mathbb{C}^2 , we show the existence of a measure of maximal dimension and identify the conditions under which a measure of full dimension exists.

1. INTRODUCTION

Let g be a hyperbolic polynomial automorphism of \mathbb{C}^2 . For $A \subset \mathbb{C}^2$ we denote by $\dim_H A$ the Hausdorff dimension of A . Let ν be an invariant Borel probability measure. We define the Hausdorff dimension of ν by

$$(1) \quad \dim_H(\nu) = \inf\{\dim_H A : \nu(A) = 1\}.$$

We define the quantity $\delta(g)$ by

$$(2) \quad \delta(g) = \sup\{\dim_H(\nu)\},$$

where the supremum is taken over all ergodic invariant Borel probability measures with positive entropy.¹ This quantity was introduced by Denker and Urbanski [DU] in the context of rational maps on the Riemann sphere. They called it the dynamical dimension of the map.

It is easy to see that the support of each measure considered in (2) is contained in the Julia set J (see Section 2 for the definition). We denote by $M(J, g|_J)$ the set of all ergodic invariant Borel probability measures supported on J .

If a measure $\nu \in M(J, g|_J)$ attains the supremum in (2), that is,

$$(3) \quad \dim_H(\nu) = \delta(g),$$

we say that ν is a measure of maximal dimension for g .

McCluskey and Manning [MM] gave a heuristic argument for the existence of a measure of maximal dimension in the case of Axiom A surface diffeomorphisms. However, it was not known until this paper whether this argument can be extended to a rigorous proof (see the remarks after Theorem 5.2 for more details).²

Received by the editors July 30, 2001 and, in revised form, December 11, 2002.

2000 *Mathematics Subject Classification*. Primary 37C45, 37D35, 32H50; Secondary 37D20, 37FXX.

The author was supported by a research fellowship of the Deutsche Forschungsgemeinschaft (DFG).

¹Using Young's formula [Y] and results by L. Barreira and the author [BW1] it is possible to show that the value of $\delta(f)$ remains the same when the supremum is taken over all invariant Borel probability measures.

²While this paper was in the process of publishing, L. Barreira and the author [BW2] proved the existence of a measure of maximal dimension for Axiom A surface diffeomorphisms.

In this paper we study the existence of a measure of maximal dimension for hyperbolic polynomial automorphisms of \mathbb{C}^2 . Our main result is the following (see Theorems 5.1, 5.2 and Corollary 5.4).

Theorem 1.1. *Let g be a hyperbolic polynomial automorphism of \mathbb{C}^2 . Then there exists a measure of maximal dimension for g . The set of measures of maximal dimension is finite.*

The proof of this theorem uses the theory of thermodynamic formalism. Other crucial ingredients are Young's formula [Y], which relates the dimension of a measure to entropy and Lyapunov exponents, and dimension-theoretic results pertaining to the dynamics of polynomial automorphisms of \mathbb{C}^2 (see [Wo1], [Wo2]). The key idea is to extract a one-parameter family of potentials and to consider the corresponding family of equilibrium measures. We show that a measure of maximal dimension necessarily belongs to this family of equilibrium measures. Furthermore, if a measure ν maximizes Hausdorff dimension among these equilibrium measures, then ν is a measure of maximal dimension.

We say that $\nu \in M(J, g|_J)$ is a measure of full dimension if

$$(4) \quad \dim_H(\nu) = \dim_H J.$$

We prove in Corollary 3.5 that there exists at most one measure of full dimension. The next result gives a classification for the existence of a measure of full dimension (see Theorems 4.1 and 4.2).

Theorem 1.2. *Let g be a hyperbolic polynomial automorphism of \mathbb{C}^2 . Then*

- i) *If g is volume-preserving, then there exists a unique measure of full dimension for g .*
- ii) *If g does not preserve volume, and if g admits a measure of full dimension, then this measure is the unique measure of maximal entropy.*

In the volume-preserving case the existence of a measure of full dimension has already been shown by Friedland and Ochs [FO]. We provide an alternative proof for this result in Theorem 4.1.

Theorem 1.2 indicates that in the case of non-volume-preserving maps, the existence of a measure of full dimension seems to be a rare phenomenon. Indeed, we show in Theorem 4.2 that the existence of such a measure is equivalent to the fact that the multipliers of all saddle points are of the same modulus (see equation (28) for the precise statement). If this condition would be satisfied, then the parameter defining the map g would provide solutions of a countable infinite set of algebraic equations. From this point of view, such an example seems very likely to not exist.

In Section 4 we observe that there exists a dense open subset of hyperbolic parameter space for which no measure of full dimension exists. This implies that

$$(5) \quad \delta(g) < \dim_H J$$

holds for these parameters.

In the last part of this paper, we analyze the dependence of $\delta(g)$ on the parameter of the mapping. More precisely, we prove the following result (see Theorems 6.1 and 6.2).

Theorem 1.3. *Let $\lambda \mapsto g_\lambda$ be a holomorphic family of hyperbolic polynomial automorphisms of \mathbb{C}^2 . Then $\lambda \mapsto \delta(g_\lambda)$ is continuous and plurisubharmonic.*

This paper is organized as follows. In Section 2 we present the basic definitions and notation, and also list the standing assumptions of the paper. In Section 3 we introduce elements from dimension theory for hyperbolic polynomial automorphisms of \mathbb{C}^2 and provide the tools for the analysis of the existence of measures of maximal and full dimension. Section 4 is devoted to the analysis of the existence of a measure of full dimension. The existence of a measure of maximal dimension is proved in Section 5. Finally, we study in Section 6 the dependence of $\delta(g)$ on parameters.

It would be interesting to understand whether, or at least under which conditions, a uniqueness result for the measure of maximal dimension holds. A partial answer to this question is given in Corollary 3.5 where a uniqueness result is shown in the case when a measure of full dimension exists.

2. NOTATION AND PRELIMINARIES

Let g be a polynomial automorphism of \mathbb{C}^2 . We denote by $\deg(g)$ the maximum of the polynomial degree of the components of g . The dynamical degree of g is defined by

$$d = \lim_{n \rightarrow \infty} (\deg(g^n))^{1/n}$$

(see [BS2], [FM]). We are interested in nontrivial dynamics, which occurs if and only if $d > 1$. Friedland and Milnor showed in [FM] that every polynomial automorphism of \mathbb{C}^2 with nontrivial dynamics is conjugate to a mapping of the form $g = g_1 \circ \dots \circ g_m$, where each g_i is a generalized Hénon mapping. This means that g_i has the form

$$(6) \quad g_i(z, w) = (w, P_i(w) + a_i z),$$

where P_i is a complex polynomial of degree $d_i \geq 2$ and a_i is a nonzero complex number. The dynamical degree d of such a map g is equal to $d_1 \cdot \dots \cdot d_m$ and therefore coincides with the polynomial degree of g .

For g we define K^\pm as the set of points in \mathbb{C}^2 with bounded forward/backward orbits, $K = K^+ \cap K^-$, $J^\pm = \partial K^\pm$ and $J = J^+ \cap J^-$. We refer to J^\pm as the positive/negative Julia set of g and call J the Julia set of g . The sets K and J are compact.

Note that the function $a = \det Dg$ is constant in \mathbb{C}^2 . Therefore, we can restrict our considerations to the volume-decreasing case ($|a| < 1$), and to the volume-preserving case ($|a| = 1$), because in the volume-increasing case ($|a| > 1$), we can consider g^{-1} .

As pointed out in the introduction we assume in this paper that g is a hyperbolic mapping. This means that there exists a continuous invariant splitting $T_J \mathbb{C}^2 = E^u \oplus E^s$ such that $Dg|_{E^u}$ is uniformly expanding and $Dg|_{E^s}$ is uniformly contracting. Hyperbolicity implies that we can associate with each point $p \in J$ its local unstable/stable manifold $W_\epsilon^{u/s}(p)$. Moreover, g is an Axiom A diffeomorphism (see [BS1] for more details).

Standing Assumptions. We now list several conditions which will be assumed throughout the entire paper:

- i) g is a polynomial automorphism of \mathbb{C}^2 with dynamical degree $d > 1$;
- ii) g is hyperbolic;
- iii) g is volume-non-increasing.

We recall that assumption iii) is actually not a restriction since we can also consider g^{-1} (see above).

3. ELEMENTS FROM DIMENSION THEORY

In this section we introduce elements from dimension theory for hyperbolic polynomial automorphisms of \mathbb{C}^2 and provide the tools for the analysis of measures of maximal and full dimension.

We start by introducing Lyapunov exponents. Let $\nu \in M(J, g|_J)$. The Julia set J is a hyperbolic set of index 1 (see [BS1]). Therefore, by the multiplicative ergodic theorem of Oseledec, there are Lyapunov exponents $\lambda_\nu^- < 0 < \lambda_\nu^+$ with respect to ν (see e.g. [KH]). In particular, ν is a hyperbolic measure. We define the quantity

$$(7) \quad \Lambda(\nu) = \lim_{n \rightarrow \infty} \frac{1}{n} \int_J \log \|Dg^n\| d\nu.$$

Similar to what was done for the measure of maximal entropy in [BS3], the positive Lyapunov exponent λ_ν^+ coincides with $\Lambda(\nu)$. Since g has constant Jacobian determinant a , the negative Lyapunov exponent λ_ν^- is given by $-\Lambda(\nu) + \log |a|$. In [Wol] we derived the formula

$$(8) \quad \Lambda(\nu) = \int_J \log \|Dg|_{E^u}\| d\nu.$$

By Young's formula [Y], we have for all $\nu \in M(J, g|_J)$ that

$$(9) \quad \dim_H(\nu) = \frac{h_\nu(g)}{\Lambda(\nu)} + \frac{h_\nu(g)}{\Lambda(\nu) - \log |a|}.$$

Here $h_\nu(g)$ denotes the measure-theoretic entropy of g with respect to ν .

Next we introduce topological pressure. Let $C(J, \mathbf{R})$ denote the Banach space of all continuous functions from J to \mathbf{R} . The topological pressure of $g|_J$, denoted by $P = P(g|_J, \cdot)$, is a mapping from $C(J, \mathbf{R})$ to \mathbf{R} (see [Wa] for the definition). The variational principle provides the formula

$$(10) \quad P(g|_J, \varphi) = \sup_{\nu \in M(J, g|_J)} \left(h_\nu(g) + \int_J \varphi d\nu \right).$$

If a measure $\nu_\varphi \in M(J, g|_J)$ achieves the supremum in equation (10), that is,

$$(11) \quad P(g|_J, \varphi) = h_{\nu_\varphi}(g) + \int_J \varphi d\nu_\varphi,$$

it is called an equilibrium measure of the potential φ .

The topological pressure has the following properties (see [R]).

- i) The topological pressure is a convex function.
- ii) If φ is a strictly negative function, then the function $t \mapsto P(g|_J, t\varphi)$ is strictly decreasing.
- iii) The topological pressure is a real analytic function on the subspace of Hölder continuous functions, that is, when $\alpha > 0$ is fixed, then $P(g|_J, \cdot)|_{C^\alpha(J, \mathbf{R})}$ is a real analytic function. Note that C^α cannot be replaced by C^0 .

- iv) If $\alpha > 0$ and $\varphi \in C^\alpha(J, \mathbf{R})$, then there exists a uniquely defined equilibrium measure $\nu_\varphi \in M(J, g|_J)$ of the potential φ . Furthermore, we have for all $\varphi, \psi \in C^\alpha(J, \mathbf{R})$,

$$(12) \quad \frac{d}{dt} \bigg|_{t=0} P(g|_J, \varphi + t\psi) = \int_J \psi d\nu_\varphi.$$

We have $P(g|_J, 0) = h_{\text{top}}(g|_J)$ (see e.g. [Wa]), where $h_{\text{top}}(g|_J)$ denotes the topological entropy of $g|_J$. Therefore, by property iv), the equilibrium measure of the potential constant zero, ν_0 , is the unique measure of maximal entropy of $g|_J$. The map g has a unique measure of maximal entropy (see [BS1]); moreover, this measure is supported on J . We conclude that ν_0 is the unique measure of maximal entropy of g . We now introduce potentials which are related to Lyapunov exponents. We define

$$(13) \quad \phi^{u/s} : J \rightarrow \mathbf{R}, \quad p \mapsto \log \|Dg(p)|_{E_p^{u/s}}\|$$

and the unstable/stable pressure functions

$$(14) \quad P^{u/s} : \mathbf{R} \rightarrow \mathbf{R}, \quad t \mapsto P(g|_J, \mp t\phi^{u/s}).$$

The Julia set J is a hyperbolic set of index 1; thus the potentials $\mp\phi^{u/s}$ are strictly negative. Therefore, property ii) of the topological pressure implies that the functions $P^{u/s}$ are strictly decreasing.

Since $\phi^{u/s}$ are Hölder continuous (see [B]), we may conclude from property iii) of the topological pressure that $P^{u/s}$ are real analytic. Property iv) of the topological pressure implies that there exist uniquely defined equilibrium measures $\nu_{\mp t\phi^{u/s}} \in M(J, g|_J)$ of the potentials $\mp t\phi^{u/s}$.

We will need the following result about the relation between the unstable and stable pressure functions.

Proposition 3.1 ([Wo2]). $P^u(t) = P^s(t) - t \log |a|$.

Lemma 3.2. $\nu_{-t\phi^u} = \nu_{t\phi^s}$.

Proof. Let $t \geq 0$. Then

$$\begin{aligned} P^s(t) &= P^u(t) + t \log |a| \\ &= h_{\nu_{-t\phi^u}}(g) + t \left(- \int \log \|Dg|_{E^u}\| d\nu_{-t\phi^u} + \log |a| \right) \\ &= h_{\nu_{-t\phi^u}}(g) + t \left(\lim_{n \rightarrow \infty} \frac{1}{n} \int - \log \|Dg^n|_{E^u}\| + \log |a^n| d\nu_{-t\phi^u} \right) \\ &= h_{\nu_{-t\phi^u}}(g) + t \left(\lim_{n \rightarrow \infty} \frac{1}{n} \int \log \|Dg^n|_{E^s}\| d\nu_{-t\phi^u} \right) \\ &= h_{\nu_{-t\phi^u}}(g) + t \int \log \|Dg|_{E^s}\| d\nu_{-t\phi^u} \\ (15) \quad &= h_{\nu_{-t\phi^u}}(g) + t \int \phi^s d\nu_{-t\phi^u}. \end{aligned}$$

The result follows from the uniqueness of the equilibrium measure of the potential $t\phi^s$. \square

We will use in the remainder of this paper the notation $\nu_t = \nu_{\mp t\phi^{u/s}}$. This notation is justified by Lemma 3.2. We also write $\Lambda(t) = \Lambda(\nu_t)$ and $h(t) = h_{\nu_t}(g)$,

and consider Λ and h as real-valued functions of t . Equations (8) and (11) imply that

$$(16) \quad P^u(t) = h(t) - t\Lambda(t).$$

Therefore, Proposition 3.1 implies that

$$(17) \quad P^s(t) = h(t) - t(\Lambda(t) - \log |a|).$$

Proposition 3.3. *Λ and h are real analytic. Furthermore,*

$$(18) \quad \frac{d\Lambda}{dt} \leq 0.$$

If Λ is not constant, then every zero of the derivative of Λ is isolated, and Λ is strictly decreasing.

Proof. Let $t_0 \geq 0$ and let ϕ^u be as in (13). We define potentials $\varphi = -t_0\phi^u$, $\psi = -\phi^u$ (here we use the notation of (12)). Therefore, application of equations (8) and (12) implies that

$$(19) \quad \frac{dP^u}{dt}(t_0) = -\Lambda(\nu_{t_0}) = -\Lambda(t_0).$$

Since P^u is real analytic, we obtain that Λ is also real analytic. We conclude from (16) that h is also real analytic. The convexity of P^u implies that

$$(20) \quad \frac{d^2P^u}{dt^2} \geq 0;$$

hence

$$(21) \quad \frac{d\Lambda}{dt} \leq 0.$$

Finally, if Λ is not constant, then the uniqueness theorem for real analytic functions, applied to the derivative of Λ , implies that all zeros of the derivative of Λ are isolated. Therefore, Λ is strictly decreasing. \square

Hausdorff dimensions of the measures ν_t . We use the notation $\Delta(t) = \dim_H(\nu_t)$. Equation (9) yields

$$(22) \quad \Delta(t) = \frac{h(t)}{\Lambda(t)} + \frac{h(t)}{\Lambda(t) - \log |a|}.$$

Thus, Δ is also a real analytic function. Equations (16), (17) and Proposition 3.1 imply that

$$(23) \quad \Delta(t) = 2t + \frac{P^u(t)}{\Lambda(t)} + \frac{P^u(t) + t \log |a|}{\Lambda(t) - \log |a|}.$$

From an elementary calculation we obtain the following formula for the derivative of Δ :

$$(24) \quad \frac{d\Delta}{dt}(t_0) = -\frac{\frac{d\Lambda}{dt}(t_0) [P^u(t_0)(\Lambda(t_0) - \log |a|)^2 + (P^u(t_0) + t_0 \log |a|)\Lambda(t_0)^2]}{\Lambda(t_0)^2(\Lambda(t_0) - \log |a|)^2}.$$

Hausdorff dimension of J . The following result due to Verjovsky and Wu provides a formula for the Hausdorff dimension of the unstable/stable slice in terms of the zeros of the pressure functions.

Theorem 3.4 ([VW]). *Let $p \in J$. Then $t^{u/s} = \dim_H W_\epsilon^{u/s}(p) \cap J$ does not depend on $p \in J$. Furthermore, $t^{u/s}$ is given by the unique solution of*

$$(25) \quad P^{u/s}(t) = 0.$$

Equation (25) is called the Bowen-Ruelle formula. We refer to $t^{u/s}$ as the Hausdorff dimension of the unstable/stable slice.

In [Wo1] we proved the formula

$$(26) \quad \dim_H J = t^u + t^s = \sup_{\nu \in M(J, g|_J)} \left(\frac{h_\nu(g)}{\Lambda(\nu)} \right) + \sup_{\nu \in M(J, g|_J)} \left(\frac{h_\nu(g)}{\Lambda(\nu) - \log |a|} \right),$$

where each of the suprema on the right-hand side of the equation is uniquely attained by the measures ν_{t^u} and ν_{t^s} , respectively. Hence

$$(27) \quad \dim_H J = \frac{h(t^u)}{\Lambda(t^u)} + \frac{h(t^s)}{\Lambda(t^s) - \log |a|}.$$

Equation (27) and the uniqueness of the measures ν_{t^u}, ν_{t^s} in equation (26) imply that, if there exists a measure of full dimension, then it already coincides with ν_{t^u} and ν_{t^s} . Thus, we have the following result.

Corollary 3.5. *Assume m is a measure of full dimension for g . Then $m = \nu_{t^u} = \nu_{t^s}$. In particular, there exists at most one measure of full dimension.*

4. MEASURES OF FULL DIMENSION

In this section we identify the conditions under which a measure of full dimension exists. We start with the volume-preserving case.

Theorem 4.1. *Let g be volume-preserving. Then $t^u = t^s$, and ν_{t^u} is a measure of full dimension for g .*

Proof. We have $|a| = 1$. Therefore Proposition 3.1 implies that $P^u = P^s$. Thus, Theorem 3.4 yields $t^u = t^s$. Therefore, by equations (9) and (27), we conclude that $\dim_H(\nu_{t^u}) = \dim_H J$, which implies that ν_{t^u} is a measure of full dimension. \square

Remark. As noted in the introduction, in the volume-preserving case, the existence of a measure of full dimension was already shown by Friedland and Ochs [FO]. They proved that the existence of a measure of full dimension follows from the fact that $|\det Dg^n(p)| = 1$ holds for every periodic point p with period n . They also observed that in this case the measure of full dimension is equivalent to the t -dimensional Hausdorff measure, where t is the Hausdorff dimension of J .

We now consider the volume-decreasing case. Let \mathcal{S} denote the set of all saddle points of g . Note that $J = \overline{\mathcal{S}}$, see [BS1]. For $p \in \mathcal{S}$ with period $n(p)$ we denote by $\lambda^{u/s}(p)$ the eigenvalues of $Dg^{n(p)}(p)$, where $|\lambda^s(p)| < 1 < |\lambda^u(p)|$. In the next theorem we provide equivalent conditions for the existence of a measure of full dimension.

Theorem 4.2. *Assume g is volume-decreasing. Then the following are equivalent.*

- i) g admits a measure of full dimension.
- ii) The unstable pressure function P^u is affine.
- iii) The stable pressure function P^s is affine.
- iv) The measure of maximal entropy is a measure of full dimension for g .

v) *The quantity*

$$(28) \quad |\lambda^u(p)|^{\frac{1}{n(p)}}$$

is independent of the periodic point $p \in \mathcal{S}$.

Proof. $ii) \Leftrightarrow iii)$ follows from Proposition 3.1.

$i) \Rightarrow ii)$ Let us assume that g admits a measure of full dimension. It is shown in [Wo2] that $t^s < t^u$. Corollary 3.5 implies that $\nu_{t^u} = \nu_{t^s}$ is the measure of full dimension. So $\Lambda(t^u) = \Lambda(t^s)$, and this in turn implies by Proposition 3.3 that the function $\Lambda(t)$ is constant. Therefore, equation (19) implies that P^u is affine.

$ii) + iii) \Rightarrow iv)$ Recall that ν_0 is the unique measure of maximal entropy for g , see Section 3. The topological entropy of $g|_J$ is equal to $\log d$ (see [BS3]). Thus $P^u(0) = P^s(0) = \log d$. Equation 19 and Proposition 3.1 imply

$$(29) \quad \frac{dP^u}{dt}(0) = -\Lambda(0),$$

$$(30) \quad \frac{dP^s}{dt}(0) = -\Lambda(0) + \log |a|.$$

Since P^u and P^s are affine, Theorem 3.4 and equation (26) imply

$$(31) \quad \dim_H J = \frac{\log d}{\Lambda(0)} + \frac{\log d}{\Lambda(0) - \log |a|}.$$

But by Young's formula (9), the right-hand side of (31) is equal to $\dim_H(\nu_0)$. Thus, ν_0 is a measure of full dimension.

$iv) \Rightarrow v)$ If ν_0 is a measure of full dimension for g , then, by Corollary 3.5, we have $\nu_0 = \nu_{t^u}$. Moreover, it follows from Theorem 3.4 that $t^u > 0$. Therefore, $v)$ follows from Proposition 4.5 of [B].

$v) \Rightarrow ii)$ follows again from Proposition 4.5 of [B].

Finally, $iv) \Rightarrow i)$ is trivial. □

Remark. As mentioned in the introduction, it is very likely that for volume-decreasing maps a measure of full dimension never exists. This can be seen by using property $v)$ of Theorem 4.2, because if such a map g exists, then the parameter defining g provides solutions of a countable infinite set of algebraic equations, which is indeed a very strong conclusion.

Let us assume that a volume-decreasing map g admits a measure of full dimension. Then it follows by equation (28) that g belongs to a real codimension one algebraic subset of parameter space. Using a perturbation argument, it is not too hard to see that we can find arbitrarily close to g a map g' for which (28) does not hold; in particular, g' has no measure of full dimension. Here we mean close with respect to the topology on hyperbolic parameter space induced by the parameter of the mapping (see [Wo1] for details). On the other hand, if g admits no measure of full dimension, then (28) does not hold for g . It can be easily shown that there is a neighborhood of g such that for each map g' in this neighborhood (28) does not hold. Therefore, there exists a dense open subset of parameters admitting no measure of full dimension. We leave the details to the reader.

5. MEASURES OF MAXIMAL DIMENSION

In this section we establish the existence of a measure of maximal dimension.

We first consider the volume-preserving case. In this situation it follows from Corollary 3.5 and Theorem 4.1 that ν_{t^u} is the unique measure of full dimension for g . By definition, every measure of full dimension is also a measure of maximal dimension. Therefore, we obtain the following.

Theorem 5.1. *Assume g is volume-preserving. Then ν_{t^u} is the unique measure of maximal dimension for g .*

We now consider the volume-decreasing case. The following theorem is the main result of this paper.

Theorem 5.2. *Assume g is volume-decreasing. Then there exists a measure of maximal dimension for g . If m is a measure of maximal dimension for g , then there exists $t^s < t < t^u$ such that m is the equilibrium measure of the potential $-\phi^u$, that is, $m = \nu_t$.*

Proof. Since g is volume-decreasing, we have $t^s < t^u$ (see [Wo2]).

We first assume that g admits a measure of full dimension. In this case application of Theorem 4.2 and Proposition 4.5 of [B] implies that ν_t does not depend on t . Moreover, ν_t is the unique measure of full and therefore also of maximal dimension. Thus, the theorem holds.

We now assume that f has no measure of full dimension. Thus, by Theorem 4.2 the functions $P^{u/s}$ are not affine. \square

Assertion 1. *There exists $\epsilon > 0$ such that Δ is strictly increasing on $[0, t^s + \epsilon)$ and strictly decreasing on $(t^u - \epsilon, \infty)$.*

Proof of Assertion 1. Theorem 3.4 and the fact that $P^{u/s}$ are strictly decreasing functions [property ii) of the topological pressure] imply that $P^s(t) > 0$ for all $t \in [0, t^s)$. Analogously we have $P^u(t) > 0$ for all $t \in [0, t^u)$. We conclude from Proposition 3.3, equation (24) and an elementary continuity argument that there exists $\epsilon > 0$ such that

$$(32) \quad \frac{d\Delta}{dt} \geq 0$$

in $[0, t^s + \epsilon)$, and all zeros of the derivative of Δ in $[0, t^s + \epsilon)$ are isolated. Therefore, Δ is strictly increasing in $[0, t^s + \epsilon)$. A similar argument shows that there exists $\epsilon > 0$ such that Δ is strictly decreasing in $(t^u - \epsilon, \infty)$.

Assertion 1 implies that there exists $t^* \in [t^s + \epsilon, t^u - \epsilon]$ such that

$$(33) \quad \dim_H(\nu_{t^*}) = \sup_{t \geq 0} \Delta(t).$$

\square

Assertion 2. *The measure ν_{t^*} is a measure of maximal dimension.*

Proof of Assertion 2. Let $(m_k)_{k \in \mathbb{N}}$ be a sequence in $M(J, g|_J)$ such that

$$(34) \quad \lim_{k \rightarrow \infty} \dim_H(m_k) = \delta(g).$$

By Assertion 1, we may assume, without loss of generality, that $\dim_H(\nu_0) = \Delta(0) < \dim_H(m_k)$ for all $k \in \mathbb{N}$. Recall that ν_0 is the unique measure of maximal entropy

of g (see Section 3). We now may conclude by Young's formula (9) that

$$(35) \qquad \Lambda(\nu_0) > \Lambda(m_k)$$

for all $k \in \mathbf{N}$. Again by Assertion 1, we may assume, without loss of generality, that $\dim_H(\nu_{t^u}) = \Delta(t^u) < \dim_H(m_k)$ for all $k \in \mathbf{N}$. Equation (26) implies that

$$(36) \qquad \frac{h_{m_k}(g)}{\Lambda(m_k)} < \frac{h(t^u)}{\Lambda(t^u)}$$

for all $k \in \mathbf{N}$. Therefore, Young's formula (9) implies that

$$(37) \qquad \frac{h_{m_k}(g)}{\Lambda(m_k) - \log |a|} > \frac{h(t^u)}{\Lambda(t^u) - \log |a|}$$

for all $k \in \mathbf{N}$. Equations (36) and (37) imply that $h_{m_k}(g) > h(t^u)$, and therefore again by equation (36) we obtain

$$(38) \qquad \Lambda(m_k) > \Lambda(t^u)$$

for all $k \in \mathbf{N}$. Since Λ is continuous, equations (35) and (38) imply that for all $k \in \mathbf{N}$ there exists $t_k \in (0, t^u)$ such that

$$(39) \qquad \Lambda(m_k) = \Lambda(t_k).$$

Thus, the variational principle (10) implies that

$$(40) \qquad h_{m_k}(g) \leq h(t_k);$$

hence

$$(41) \qquad \dim_H(m_k) \leq \Delta(t_k)$$

for all $k \in \mathbf{N}$. This implies that

$$(42) \qquad \dim_H(m_k) \leq \dim_H(\nu_{t^*})$$

for all $k \in \mathbf{N}$. We conclude that ν_{t^*} is a measure of maximal dimension.

To complete the proof of the theorem we have to show the following.

Assertion 3. *For every measure m of maximal dimension there exists $t^s < t < t^u$ such that m is the equilibrium measure of the potential $-t\phi^u$.*

Proof of Assertion 3. Let m be a measure of maximal dimension. We apply to m (instead of m_k) the same argument as in the proof of Assertion 2. This implies that there exists $t \in (0, t^u)$ such that $\Lambda(m) = \Lambda(t)$. Since $\dim_H(m) \geq \Delta(t)$, we may deduce by equation (9) that $h_m(g) \geq h(t)$. On the other hand, since ν_t is the equilibrium measure of the potential $-t\phi^u$, we may conclude by (10) and (11) that $h_m(g) \leq h(t)$. Hence $h_m(g) = h(t)$. Therefore, the uniqueness of the equilibrium measure of the potential $-t\phi^u$ implies that $m = \nu_t$. Finally, Assertion 1 implies that $t \in (t^s, t^u)$. This completes the proof. \square

Remarks. The following heuristic argument was given by McCluskey and Manning [MM] to state the existence of a measure of maximal dimension in the case of C^2 axiom A diffeomorphisms of real surfaces. Since the entropy map is upper semi-continuous, it can be shown that the map $\nu \mapsto \dim_H(\nu)$, defined on the set of all ergodic invariant measures, is also upper semi-continuous. It is now suggested in [MM] that this implies the existence of a measure of maximal dimension. To make this argument rigorous we need to show that there exists a sequence of ergodic invariant measures m_k with $\dim_H(m_k) \rightarrow \delta(g)$ having an ergodic weak* limit.

Whether this holds is not clear since the set of all ergodic invariant measures is dense in the set of all invariant measures with respect to the weak* topology; see Proposition 21.9 in [DGS]. In particular, the set of all ergodic invariant measures is not closed.

In the following we describe properties of the measures of maximal dimension.

Corollary 5.3. *Assume that g admits no measure of full dimension. Let $t \geq 0$ be such that ν_t is a measure of maximal dimension. Then*

$$(43) \quad \delta(g) = 2t + \frac{P^u(t) \log |a|}{\Lambda(t)^2}.$$

Proof. By Proposition 3.3, equation (24) and Theorem 5.2, a necessary condition for ν_t being a measure of maximal dimension is

$$(44) \quad P^u(t)(\Lambda(t) - \log |a|)^2 + (P^u(t) + t \log |a|)\Lambda(t)^2 = 0.$$

Therefore, the result follows from equation (23). \square

Corollary 5.4. *The set of all measures of maximal dimension is finite.*

Proof. Assume first that g admits a measure of full dimension. Then, this measure is the unique measure of maximal dimension. If g has no measure of full dimension, then the function Δ is a non-constant real analytic function on $[0, \infty)$. Therefore, it follows from the uniqueness theorem for real analytic functions that Δ has only finitely many maxima in $[t^s, t^u]$. The result follows from Theorem 5.2. \square

Corollary 5.5. *Every measure ν of maximal dimension is Bernoulli.*

Proof. Since $g|_J$ is topological mixing (see [BS1]), the result follows from the fact that ν is an equilibrium measure of a Hölder continuous potential (see [B], Thm. 4.1). \square

6. DEPENDENCE ON PARAMETERS

Let A denote an open subset of \mathbb{C}^k and let $(g_\lambda)_{\lambda \in A}$ be a holomorphic family of hyperbolic polynomial automorphisms of \mathbb{C}^2 of fixed dynamical degree $d > 1$. We denote by J_λ the Julia set, by a_λ the Jacobian determinant, and by $P_\lambda^{u/s}$ the unstable/stable pressure functions of g_λ . We also write $\Delta_\lambda(t)$ instead of $\Delta(t)$. First, we show that $\delta(g)$ depends continuously on the parameter of the mapping.

Theorem 6.1. *The function $\lambda \mapsto \delta(g_\lambda)$ is continuous in A .*

Proof. Let $\lambda_0 \in A$. The result of [VW] implies that there exist $\epsilon > 0$ and a real analytic function

$$(45) \quad \mathcal{P} : B(\lambda_0, \epsilon) \times [0, \infty) \rightarrow \mathbf{R},$$

such that $\mathcal{P}(\lambda, \cdot) = P_\lambda^u$ for all $\lambda \in B(\lambda_0, \epsilon)$. Therefore, equations (19) and (23) imply that

$$(46) \quad \mathcal{D} : B(\lambda_0, \epsilon) \times [0, \infty) \rightarrow \mathbf{R}, \quad (\lambda, t) \mapsto \Delta_\lambda(t)$$

is also a real analytic function. Now we may conclude by Theorem 5.1 and Theorem 5.2 that

$$(47) \quad \delta(g_\lambda) = \max_{t \in [0, 2]} \mathcal{D}(\lambda, t).$$

The result follows by an elementary continuity argument. \square

Remark. McCluskey and Manning [MM] considered C^2 Axiom A diffeomorphisms of real surfaces. They showed that for these mappings $\delta(g)$ depends continuously on the mapping with respect to the C^2 topology.

Finally, we show that $\delta(g)$ depends plurisubharmonically on the parameter of the mapping.

Theorem 6.2. *The function $\lambda \mapsto \delta(g_\lambda)$ is plurisubharmonic in A .*

Proof. Let $g_0 \in A$ and let L be a complex line in \mathbb{C}^k containing g_0 . Then there exists a holomorphic family $(g_\lambda)_{\lambda \in D}$, where D is a disk with center 0 in \mathbb{C} such that $\{g_\lambda : \lambda \in D\}$ is a neighborhood of g_0 in $L \cap A$. If the radius of D is small enough, then there exists a family $(\kappa_\lambda)_{\lambda \in D}$, where each κ_λ is the uniquely defined conjugacy between $g_0|_{J_0}$ and $g_\lambda|_{J_\lambda}$. Therefore, $T_\lambda = (\kappa_\lambda)_*$ defines a family of bijections from $M(J_0, g_0|_{J_0})$ to $M(J_\lambda, g_\lambda|_{J_\lambda})$. Moreover, we have $h_\nu(g_0) = h_{T_\lambda(\nu)}(g_\lambda)$ for all $\nu \in M(J_0, g_0|_{J_0})$ and all $\lambda \in D$ (see [Wol] for the details). In [Wol] we showed that if $\nu \in M(J_0, g_0|_{J_0})$ is fixed, then $\lambda \mapsto \Lambda(T_\lambda(\nu))$ is a harmonic function in D . We conclude by Young's formula (9) that

$$(48) \quad \delta(g_\lambda) = \sup_{\nu \in M(J_0, g_0|_{J_0})} \left(\frac{h_\nu(g_0)}{\Lambda(T_\lambda(\nu))} + \frac{h_\nu(g_0)}{\Lambda(T_\lambda(\nu)) - \log |a_\lambda|} \right).$$

The functions $\lambda \mapsto \Lambda(T_\lambda(\nu))$, $\lambda \mapsto \Lambda(T_\lambda(\nu)) - \log |a_\lambda|$ are harmonic in D . Note that $x \mapsto x^{-1}$ is a convex function on \mathbf{R}^+ . This implies that the functions $\lambda \mapsto \Lambda(T_\lambda(\nu))^{-1}$, $\lambda \mapsto (\Lambda(T_\lambda(\nu)) - \log |a_\lambda|)^{-1}$ are subharmonic in D . The continuous function $\lambda \mapsto \delta(g_\lambda)$ is therefore given by the supremum over a family of subharmonic functions. We conclude that the function $\lambda \mapsto \delta(g_\lambda)$ is subharmonic in D . This completes the proof. \square

ACKNOWLEDGEMENT

I would like to thank Eric Bedford and Marlies Gerber for many useful discussions and comments during the preparation of this paper. I also would like to thank the referee for his detailed report, and for suggesting various improvements. This paper was written while I was visiting Indiana University and I would like to thank the Department of Mathematics for its warm hospitality.

REFERENCES

- [B] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, *Lecture Notes in Math.* **470**, Springer-Verlag, Berlin, 1975 MR **56**:1364
- [BS1] E. Bedford and J. Smillie, Polynomial diffeomorphisms of \mathbb{C}^2 : Currents, equilibrium measure and hyperbolicity, *Invent. Math.* **103** (1991), 69 - 99 MR **92a**:32035
- [BS2] E. Bedford and J. Smillie, Polynomial diffeomorphisms of \mathbb{C}^2 2: Stable manifolds and recurrence, *J. Amer. Math. Soc.* **4** (1991), 657 - 679 MR **92m**:32048
- [BS3] E. Bedford and J. Smillie, Polynomial diffeomorphisms of \mathbb{C}^2 3: Ergodicity, exponents and entropy of the equilibrium measure, *Math. Ann.* **294** (1992), 395 - 420 MR **93k**:32062
- [BW1] L. Barreira and C. Wolf, Pointwise dimension and ergodic decompositions, preprint, 2002
- [BW2] L. Barreira and C. Wolf, Measures of maximal dimension for hyperbolic diffeomorphisms, *Comm. Math. Phys.*, to appear
- [DGS] M. Denker, C. Grillenberger and K. Sigmund, Ergodic theory on compact spaces, *Lecture Notes in Math.* **527**, Springer-Verlag, Berlin, 1976 MR **56**:15879
- [DU] M. Denker and M. Urbanski, On Sullivan's conformal measures for rational maps on the Riemann sphere, *Nonlinearity* **4** (1991), 365-384 MR **92f**:58097
- [FM] S. Friedland and J. Milnor, Dynamical properties of plane polynomial automorphisms, *Ergodic Theory and Dynam. Syst.* **9** (1989), 67-99 MR **90f**:58163

- [FO] S. Friedland and G. Ochs, Hausdorff dimension, strong hyperbolicity and complex dynamics, *Discrete and Continuous Dynam. Syst.* **4** (1998), 405 - 430 MR **99g**:58091
- [KH] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, 1995 MR **96c**:58055
- [MM] H. McCluskey and A. Manning, Hausdorff dimension for horseshoes, *Ergodic Theory and Dynam. Syst.* **3** (1983), 251 - 260 MR **85j**:58127
- [R] D. Ruelle, *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics*, Addison-Wesley, Reading, MA, 1978 MR **80g**:82017
- [VW] A. Verjovsky and H. Wu, Hausdorff dimension of Julia sets of complex Hénon mappings, *Ergodic Theory and Dynam. Syst.* **16** (1996), 849 - 861 MR **97g**:58143
- [Wa] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer, 1982 MR **84e**:28017
- [Wo1] C. Wolf, Dimension of Julia sets of polynomial automorphisms of \mathbb{C}^2 , *Michigan Mathematical Journal* **47** (2000), 585 - 600 MR **2002a**:37072
- [Wo2] C. Wolf, Hausdorff and topological dimension for polynomial automorphisms of \mathbb{C}^2 , *Ergodic Theory and Dynam. Syst.* **22** (2002), 1313-1327
- [Y] L.-S. Young, Dimension, entropy and Lyapunov exponents, *Ergodic Theory and Dynam. Syst.* **2** (1982), 109-124 MR **84h**:58087

DEPARTMENT OF MATHEMATICS, WICHITA STATE UNIVERSITY, WICHITA, KANSAS 67260-0033

E-mail address: cwolf@math.twsu.edu

URL: <http://www.math.wichita.edu/~cwolf/>

HAUSDORFF DIMENSION AND ASYMPTOTIC CYCLES

MARK POLLICOTT

ABSTRACT. We carry out a multifractal analysis for the asymptotic cycles for compact Riemann surfaces of genus $g \geq 2$. This describes the set of unit tangent vectors for which the associated orbit has a given asymptotic cycle in homology.

0. INTRODUCTION

For many years Hausdorff dimension has played an important role in understanding some subtle features in both geometry and dynamical systems. In this note we use it as a tool in studying geodesic flows and the homology of surfaces. Let V be a compact oriented Riemann surface with genus $g \geq 2$ and consider the geodesic flow $\phi_t : S_1 V \rightarrow S_1 V$ on the unit tangent bundle. We can associate to each ϕ -invariant probability measure μ a Schwartzman asymptotic cycle $\Lambda_\mu \in H^1(V, \mathbb{R})^*$. By Poincaré duality we can identify $H_1(V, \mathbb{R}) = H^1(V, \mathbb{R})^*$ and we can trivially identify $H_1(V, \mathbb{R}) = \mathbb{R}^{2g}$.

There is a particularly simple geometric way to understand Λ_μ (cf. [Su]). Given $v \in S_1 M$ and large T we can close up the orbit segment $\phi_{[0, T]} v$ by an arc of uniformly bounded length to get a closed curve $\gamma_{v, T}$, say. Then for almost every v (with respect to μ) we have that $[\gamma_{v, T}]/T \rightarrow \Lambda_\mu$ as $T \rightarrow \infty$. The range of values $\mathcal{B} \subset H_1(V, \mathbb{R})$ that can occur for different invariant measures is the “unit ball in the stable norm”. We shall prove the following result.

Theorem 1. *For $\underline{\alpha} \in \text{int}(\mathcal{B})$ in the interior of \mathcal{B} , the set of unit tangent vectors $X_{\underline{\alpha}} \subset S_1 V$ for which the limit exists and equals α is dense and uncountable. Moreover, the map $\text{int}(\mathcal{B}) \ni \underline{\alpha} \rightarrow \dim_H(X_{\underline{\alpha}})$ is analytic.*

Our approach uses a multivariable multifractal result on interval maps (Lemma 5), for which we give a simple proof based on the elegant argument in [PW]. Previously, Pesin and Sadovskaya studied multifractal analysis for Anosov flows [PS], but the exact type of multivariable results we require are not covered there. On the other hand, Barreira, Saussol and Schmeling have developed a general theory of multivariable multifractal analysis [BSS], but their results are formulated from the point of view of entropies, rather than Hausdorff dimension (cf. [Je2]).

Received by the editors October 25, 2002.

2000 *Mathematics Subject Classification.* Primary 28A78, 37D35; Secondary 37D40, 55N10.

I am very grateful to Howie Weiss and Luis Barreira for very useful conversations on multifractal analysis.

1. ASYMPTOTIC CYCLES AND THE SET \mathcal{B}

We begin with some general background and useful results on asymptotic cycles. The following definition was introduced by Schwartzman in 1957 [Sc].

Definition. To each ϕ -invariant measure μ we associate a linear functional $\Lambda_\mu \in H^1(V, \mathbb{R})^*$ ($= H_1(V, \mathbb{R})$) by

$$\Lambda_\mu(\omega) = \int \omega(v) d\mu(v),$$

where ω is a closed 1-form representing an element $[\omega] \in H^1(V, \mathbb{R})$ in the de Rham cohomology. We call Λ_μ the *asymptotic cycle* for μ .

The following result is well known.

Lemma 1 [Sc]. *If λ is the Liouville measure, then $\Lambda_\lambda = 0$.*

Proof. Since the proof is simple, we include it for completeness. Geodesic flows have a natural involution $i : S_1 V \rightarrow S_1 V$ so that the image $i(v)$ of a unit tangent vector v has the same base point, but points in the opposite direction. Since $i_* \lambda = \lambda$, it is easy to see that $\Lambda_\lambda(\omega) = \int \omega(v) d\lambda(v) = 0$ for all closed 1-forms ω . \square

For other simple examples of Anosov Hamiltonian flows the Liouville measure will still be preserved, but the corresponding asymptotic cycle may well be nonzero.

Definition. The Federer *stable norm* $\|\cdot\|$ on $H_1(V, \mathbb{R})$ is defined by

$$\|v\| = \inf \left\{ \sum_i |r_i| \cdot \text{length}(\gamma_i) : v = \sum_i r_i \gamma_i \text{ with } r_i \in \mathbb{R}, \gamma_i = \text{closed curve} \right\}.$$

Let $\mathcal{B} \subset H_1(V, \mathbb{R})$ denote the closed unit ball with respect to the stable norm.

A well-known classical conjecture (related to the Hopf conjecture cf. [BI]) states that the stable norm for a torus is Euclidean if and only if the metric is flat. Bangert showed that the stable norm of a 2-torus is differentiable in a rational direction if and only if the corresponding minimizing geodesics foliate the torus [Ba]. The set \mathcal{B} was studied for surfaces of higher genus by McShane and Rivin [MR] and Massart [Ma1], [Ma2]. McShane and Rivin showed that the stable norm for a once-punctured torus is never differentiable in a rational direction, and Massart extended this to compact oriented Riemann surfaces V with genus $g \geq 2$.

Lemma 2 [Ma2]. *The set \mathcal{B} is closed and convex. However, \mathcal{B} is not strictly convex and the boundary is not differentiable in rational directions.*

Let us consider a more dynamical viewpoint. For each (directed) closed geodesic γ we can associate its real homology class $[\gamma] \in H_1(V, \mathbb{R})$ and its length $l(\gamma)$. Two equivalent presentations of $\mathcal{B} \subset H_1(V, \mathbb{R})$ are (cf. [Ma1])

$$\begin{aligned} \mathcal{B} &= \{ \Lambda_\mu : \mu = \phi - \text{invariant probability} \} \\ &= \overline{\left\{ \frac{[\gamma]}{l(\gamma)} : \gamma = \text{is a closed geodesic} \right\}}. \end{aligned}$$

Observe that by fixing a suitable basis of (harmonic) 1-forms $\omega_1, \dots, \omega_n$ for $H^1(V, \mathbb{R})$ we can write $\Lambda_\mu = \alpha_1 \omega_1 + \dots + \alpha_{2g} \omega_{2g}$ and therefore represent Λ_μ in coordinates

as $\underline{\alpha} = (\alpha_1, \dots, \alpha_{2g})$. We can now define functions $f_i : S_1V \rightarrow \mathbb{R}$ ($i = 1, \dots, 2g$) on the unit tangent bundle by $f_i(v) = \omega_i(v)$ and rewrite

$$X_{\underline{\alpha}} = \left\{ v \in S_1V : \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f_j(\phi_t v) dt = \alpha_j, \text{ for } j = 1, \dots, 2g \right\}.$$

We briefly recall the definition of Hausdorff dimension. Let (Y, d) be a metric space. Given $\epsilon, \delta > 0$ we can define

$$H_\delta^s(Y) = \inf \left\{ \sum_{U \in \mathcal{U}} \text{diam}(U)^s : \mathcal{U} = \delta\text{-cover for } Y \right\},$$

where the summation is over all covers \mathcal{U} consisting of open sets of diameter at most δ . We can then let $H^s(Y) = \lim_{\delta \rightarrow 0} H_\delta^s(Y)$ and define the *Hausdorff dimension* of Y by $\dim_H(Y) = \inf \{ \delta > 0 : H^\delta(Y) = 0 \}$. We can write S_1V as a disjoint union

$$S_1V = X_0 \cup \left(\bigcup_{\underline{\alpha} \in H_1(V, \mathbb{R}) - \{0\}} X_{\underline{\alpha}} \right) \cup X_\infty,$$

where

- (a) $X_0 = \{v \in X : \lim_{T \rightarrow +\infty} [\gamma_{v,T}]/T = 0\}$ has full Liouville measure;
- (b) $X_{\underline{\alpha}} = \{v \in X : \lim_{T \rightarrow +\infty} [\gamma_{v,T}]/T = \underline{\alpha}\}$ has zero Liouville measure for $\underline{\alpha} \neq 0$;
- (c) X_∞ is the set of points for which $[\gamma_{v,T}]/T$ does not converge.

We recall that a ϕ -invariant probability measure μ is a *Gibbs measure* (for a Hölder continuous function $F : S_1V \rightarrow \mathbb{R}$, say) if

$$h(\phi, \mu) + \int F d\mu \geq h(\phi, \nu) + \int F d\nu$$

for all ϕ -invariant probability measures ν . The following more precise result subsumes Theorem 1.

Proposition 1.

- (1) The Hausdorff dimension $\dim_H(X_{\underline{\alpha}})$ satisfies $\dim_H(X_{\underline{\alpha}}) \leq 3$, with equality if and only if $\underline{\alpha} = 0$;
- (2) The Hausdorff dimension $\dim_H(X_{\underline{\alpha}})$ depends real analytically on $\underline{\alpha} \in \text{int}(\mathcal{B})$;
- (3) For $\underline{\alpha} \in \text{int}(\mathcal{B})$, the set $X_{\underline{\alpha}} \subset S_1V$ is dense and uncountable and carries a Gibbs measure.

Remark. By contrast, if $\underline{\alpha}$ is a point in the boundary of \mathcal{B} , then $\Lambda_{\underline{\alpha}}$ may consist of a single orbit (corresponding to a simple closed geodesic) [Ma1], [Ma2].

To prove Proposition 1 it is convenient to reduce the problem to one for interval maps. In particular, the following standard result helps to relate the flow to a simple one-dimensional system.

Lemma 3. *There exists an expanding C^ω Markov interval map $T : I \rightarrow I$ and locally constant functions $\psi_i : I \rightarrow \mathbb{R}$ ($i = 1, \dots, 2g$) such that:*

- (1) the prime closed ϕ -orbits γ correspond to prime periodic $T : I \rightarrow I$ orbits $T^n x = x$ (with at most a finite number of exceptions);
- (2) the least period of γ is given by $l(\gamma) = \log |(T^n)'(x)|$;

- (3) the homology class of γ is given by $[\gamma] = (\psi_1^n(x), \dots, \psi_{2g}^n(x))$, where we write $\psi_i^n(x) = \psi(x) + \psi_i(Tx) + \dots + \psi_i(T^{n-1}x)$, for $1 \leq i \leq 2g$ and $n \geq 1$.

Proof. The construction is fairly standard; so we shall only give an outline. We begin by choosing (two-dimensional) Markov Poincaré sections T_1, \dots, T_k to the geodesic flow $\phi_t : S_1V \rightarrow S_1V$ [Ra], [Bw1]. Moreover, we can assume that the sections are foliated by stable manifolds. In particular, we can identify the sections along the stable manifolds so that each section T_i quotients to an interval I_i , say. By virtue of the Markov property of the sections, the Poincaré (first return) map on $\bigcup_{i=1}^n T_i$ induces a Markov interval map $T : I \rightarrow I$, where $I = \bigcup_{i=1}^n I_i$. Since the foliation of S_1V by stable manifolds is analytic, we can deduce that $T : I \rightarrow I$ is C^ω . There is a one-to-one correspondence between closed ϕ -orbits γ , periodic orbits for the Poincaré map and T -periodic points, except possibly for the (at most) finite number of closed ϕ -orbits that pass through the boundary of a section.

For part(2), it is an easy observation that since the surface has constant curvature $\kappa = -1$, say, the length $l(\gamma)$ is equal to the expansion coefficient (or Lyapunov exponent) in the unstable direction around the orbit. If γ corresponds to the periodic orbit $T^n x = x$, then the associated expansion coefficient around the T -orbit is precisely $\log |(T^n)'(x)|$.

For part (3), we follow a construction of Franks [Fr]. We can choose a fixed reference point $x_0 \in V$ and continuous paths $\rho_i : [0, 1] \rightarrow V$ ($i = 1, \dots, k$) such that $\rho_i(0) = x_0$ and $\rho_i(1)$ is contained in the image $\pi(T_i)$ under the canonical projection $\pi : S_1V \rightarrow V$. Assuming that $T^{-1}(\text{int}(I_i)) \cap \text{int}(I_j) \neq \emptyset$, we can associate a closed curve c_{ij} consisting of the composition of ρ_i ; a curve approximating the geodesic arc from $\pi(T_i)$ to $\pi(T_j)$; and ρ_j^{-1} . We let $[c_{ij}] \in H_1(V, \mathbb{R}) = \mathbb{R}^{2g}$ be the associated element in homology. We can then define $\psi_i : I \rightarrow \mathbb{R}$ by the coordinate functions, i.e., $(\psi_1(x), \dots, \psi_{2g}(x)) = [c_{ij}]$ when $x \in \text{int}(I_i)$ and $Tx \in \text{int}(I_j)$. In particular, we see that if $T^j x \in I_{i_j}$, for $j = 0, \dots, n-1$, then

$$\begin{aligned} [\gamma] &= [c_{i_0 i_1}] \circ \dots \circ [c_{i_{n-2} i_{n-1}}] \circ [c_{i_{n-1} i_0}] \\ &= (\psi_1^n(x), \dots, \psi_{2g}^n(x)). \end{aligned}$$

Moreover, the functions ψ_i are constant on $T^{-1}(\text{int}(I_j)) \cap \text{int}(I_i)$. □

To prove Proposition 1, we shall want to apply the following general result on Markov interval maps.

Lemma 4. Let $T : I \rightarrow I$ be an expanding $C^{1+\beta}$ Markov interval map, for some $0 < \beta \leq 1$, and let $\psi_i : I \rightarrow \mathbb{R}$ (for $i = 1, \dots, N$) be C^β functions. Let $\underline{\alpha} = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$. Let

$$Y_{\underline{\alpha}}(T, I) = \left\{ \omega \in I : \lim_{n \rightarrow +\infty} \frac{\psi_j^n(\omega)}{\log |(T^n)'(\omega)|} = \alpha_j, \text{ for } j = 1, \dots, N \right\}.$$

Let $\mathcal{B} \subset \mathbb{R}^N$ be the range of $\left(\frac{\int \psi_1 dm}{\int \log |T'| dm}, \dots, \frac{\int \psi_N dm}{\int \log |T'| dm} \right)$, where m ranges over all T -invariant probability measures. For $\underline{\alpha} \in \text{int}(\mathcal{B})$ we have that:

- (1) The Hausdorff dimension $\dim_H(Y_{\underline{\alpha}}(T, I))$ satisfies $\dim_H(Y_{\underline{\alpha}}(T, I)) \leq 1$, with equality if and only if $\underline{\alpha} = 0$;
- (2) The Hausdorff dimension $\dim_H(Y_{\underline{\alpha}}(T, I))$ depends real analytically on $\underline{\alpha}$ on the interior of \mathcal{B} ;

- (3) For $\underline{\alpha} \in \text{int}(\mathcal{B})$, the set $Y_{\underline{\alpha}}(T, I) \subset I$ is dense, uncountable and carries a Gibbs measure.

Lemma 4 can be deduced as a special case of Theorems 8 and 13 in [BSS] (and observation 2 after Theorem 8). However, we can present a self-contained independent proof of this lemma in this paper.

Lemma 5. *We can identify $\dim_H(X_{\underline{\alpha}}) = \dim_H(Y_{\underline{\alpha}}(T, I)) + 2$.*

Proof. It is immediate from the definitions that $X_{\underline{\alpha}}$ is ϕ -invariant. In particular, we can write that $\dim_H(X_{\underline{\alpha}}) = \dim_H(X_{\underline{\alpha}} \cap (\bigcup_{i=1}^k T_i)) + 1 = \max_{1 \leq i \leq k} \dim_H(X_{\underline{\alpha}} \cap T_i) + 1$. Moreover, if $v \in X_{\underline{\alpha}} \cap T_i$, for some $1 \leq i \leq k$, then all points lie on the same piece of stable manifold in T_i as v . Since the foliation of each section T_i is Lipschitz (even C^ω), we deduce that $\dim_H(X_{\underline{\alpha}} \cap T_i) = \dim_H(Y_{\underline{\alpha}}(T, I)) + 1$. This completes the proof. \square

Proposition 1 now follows from Lemmas 4 and 5. A slight modification of this analysis gives the following stronger result.

Proposition 2. *Given $\underline{\alpha}, \underline{\beta} \in H_1(V, \mathbb{R})$, let $X_{\underline{\alpha}, \underline{\beta}} \subset S_V$ be the set of unit tangent vectors v such that $\lim_{T \rightarrow \infty} [\gamma_{v, T}]/T = \underline{\alpha}$ and $\lim_{T \rightarrow \infty} [\gamma_{v, -T}]/T = \underline{\beta}$ (by which we denote the asymptotic cycle flowing backwards in time).*

- (1) *The Hausdorff dimension $\dim_H(X_{\underline{\alpha}, \underline{\beta}})$ satisfies $\dim_H(X_{\underline{\alpha}, \underline{\beta}}) \leq 3$, with equality if and only if $\underline{\alpha} = \underline{\beta} = 0$;*
- (2) *The Hausdorff dimension $\dim_H(X_{\underline{\alpha}, \underline{\beta}})$ depends real analytically on $\underline{\alpha}, \underline{\beta} \in \text{int}(\mathcal{B})$;*
- (3) *For $\underline{\alpha}, \underline{\beta} \in \text{int}(\mathcal{B})$, the set $X_{\underline{\alpha}, \underline{\beta}} \subset S_1 V$ is dense and uncountable and carries a Gibbs measure.*

Proof. If we had chosen in the proof of Lemma 3 to change the direction of the flow, i.e., replace the flow ϕ_t by the flow ϕ_{-t} , then this would have had the effect of reversing the direction of the Poincaré map and interchanging the stable and unstable manifolds. Following the same steps in the construction as before we would arrive at another C^ω expanding Markov map $S : J \rightarrow J$, say. By a similar reasoning to that above we can deduce that $X_{\underline{\alpha}, \underline{\beta}} \cap T_i$ is locally diffeomorphic to $Y_{\underline{\alpha}}(T, I) \times Y_{\underline{\beta}}(S, J)$, for each $1 \leq i \leq k$. Moreover, we can write $\dim_H(X_{\underline{\alpha}, \underline{\beta}}) = \dim_H(Y_{\underline{\alpha}}(T, I)) + \dim_H(Y_{\underline{\beta}}(S, J)) + 1$. The result then follows from Lemma 4. \square

2. SUBSHIFTS AND GIBBS MEASURES

We now need to do some preliminary work in order to prove Lemma 4. The expanding Markov interval map $T : I \rightarrow I$ can be modelled by a one-sided subshift of finite type. More precisely, we define the $k \times k$ matrix A by $A(i, j) = 1$ if $T^{-1}\text{int}(T_j) \cap \text{int}(T_i) \neq \emptyset$, and 0 otherwise. We can define

$$\Sigma_A^+ = \left\{ x \in \prod_{n=0}^{\infty} \{1, \dots, k\} : A(x_n, x_{n+1}) = 1, n \geq 0 \right\}$$

and a metric $d(x, y) = \sum_{n=0}^{\infty} 2^{-n}(1 - \delta(x_n, y_n))$. We define a local homeomorphism $\sigma : \Sigma_A^+ \rightarrow \Sigma_A^+$ by $(\sigma x)_n = x_{n+1}$. The Hölder map $\pi : \Sigma_A^+ \rightarrow I$ defined by $\pi(x) = \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$ is a semi-conjugacy i.e., $\pi \sigma = T \pi$. We define Hölder functions $\hat{\psi}_j : \Sigma_A^+ \rightarrow \mathbb{R}$, for $j = 1, \dots, 2g$ by $\hat{\psi}_j = \psi_j \circ \pi$.

To proceed we need to impose the following hypothesis on the functions.

Hypothesis. There are no solutions $b_1, \dots, b_{2g} \in \mathbb{R}$ and $u \in C(\Sigma_A^+, \mathbb{R})$ to

$$\log |T'| \circ \pi + \sum_{j=1}^{2g} b_j \widehat{\psi}_j + u \circ \sigma - u = 0.$$

It is easy to check that this hypothesis holds in our application by considering closed orbits. In particular, the above identity would imply that the lengths of closed geodesics lie in the semi-group $b_1 \mathbb{Z}^+ + \dots + b_{2g} \mathbb{Z}^+$, giving an obvious contradiction.

We define the *pressure* map $P : C^\gamma(\Sigma_A^+, \mathbb{R}) \rightarrow \mathbb{R}$ by

$$P(g) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \left(\sum_{\sigma^n x = x} \exp(g^n(x)) \right)$$

on Hölder continuous functions. This map is known to be real analytic. Furthermore, the hypothesis allows us to deduce that the map $t \mapsto P(-t \log |T'| \circ \pi + q_1 \widehat{\psi}_1 + \dots + q_{2g} \widehat{\psi}_{2g})$ is strictly convex [Ru]. We can now define the following.

Definition. Given $\underline{q} = (q_1, \dots, q_{2g}) \in \mathbb{R}^{2g}$, we let $t = t(\underline{q}) \in \mathbb{R}$ be the (unique) solution to

$$P \left(-t \log |T'| \circ \pi + q_1 \widehat{\psi}_1 + \dots + q_{2g} \widehat{\psi}_{2g} \right) = 0.$$

Remark. Since $\log |T'| > 0$ we see that for fixed \underline{q} the map $t \mapsto P(-t \log |T'| \circ \pi + q_1 \widehat{\psi}_1 + \dots + q_{2g} \widehat{\psi}_{2g})$ is a bijection on \mathbb{R} .

The analyticity of the map $P : C^\gamma(\Sigma_A^+) \rightarrow \mathbb{R}$ and the implicit function theory (using the estimate on the derivative in the next lemma) allow us to deduce that $t(\underline{q})$ has a real analytic dependence on \underline{q} .

We also need to consider the associated Gibbs measure. Let us define a cylinder set in Σ_A^+ by $[x_0, \dots, x_{n-1}] = \{y \in \Sigma_A^+ : x_i = y_i, \text{ for } 0 \leq i \leq n-1\}$. We have the following useful characterization.

Lemma 6 (cf. [Bw2]). *There exists $C > 0$ and a σ -invariant probability measure $\mu = \mu_{t, \underline{q}}$ (called a Gibbs measure) such that*

$$(2.1) \quad \frac{1}{C} \leq \frac{\mu[x_0, \dots, x_{n-1}]}{|(T^n)'(\pi x)|^{-t(\underline{q})} e^{q_1 \widehat{\psi}_1^n(x) + \dots + q_{2g} \widehat{\psi}_{2g}^n(x)}} \leq C, \quad \forall x \in \Sigma_A^+, \quad \forall n \geq 1,$$

where we denote $\widehat{\psi}_i^n(x) = \widehat{\psi}_i(x) + \widehat{\psi}_i(\sigma x) + \dots + \widehat{\psi}_i(\sigma^{n-1} x)$, for $1 \leq i \leq 2g$.

The next lemma shows that given $\underline{\alpha} \in \text{int}(\mathcal{B})$, we can associate a Gibbs measure μ .

Lemma 7. *Given $\underline{\alpha} \in \text{int}(\mathcal{B})$, we can find μ such that*

$$(2.2) \quad \alpha_j := \frac{\int \widehat{\psi}_j d\mu}{\int \log |T'| \circ \pi d\mu}, \text{ for } j = 1, \dots, 2g.$$

Proof. Given $\underline{\alpha} \in \text{int}(\mathcal{B})$, we can define

$$h(\underline{\alpha}) := \sup \left\{ \frac{h(m)}{\int \log |T'(x)| dm} : m = T\text{-invariant probability with } \frac{\int \widehat{\Psi} dm}{\int \log |T'(x)| dm} = \underline{\alpha} \right\},$$

where $\widehat{\Psi} = (\widehat{\psi}_1, \dots, \widehat{\psi}_{2g})$. In particular, given any $\epsilon > 0$ we can choose m such that $h(m) \geq (h(\underline{\alpha}) - \epsilon) \int \log |T'| dm$ and then by the variational principle:

$$\begin{aligned} & P(-h(\underline{\alpha}) \log |T'| + \sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|)) \\ & \geq h(m) + \int \left(-h(\underline{\alpha}) \log |T'| + \sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|) \right) dm \\ & \geq -\epsilon \int \log |T'| dm + \underbrace{\int \left(\sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|) \right) dm}_{=0}. \end{aligned}$$

Since ϵ is arbitrary, we deduce that

$$(2.3) \quad P(-h(\underline{\alpha}) \log |T'| + \sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|)) \geq 0,$$

for all $\underline{q} \in \mathbb{R}^{2g}$. Considering (2.3) with $\underline{q}/\|\underline{q}\|_2 \in \mathbb{R}^{2g}$ replacing \underline{q} we can deduce from the variational principle that there exists an invariant probability measure m , say, with

$$\int \left(\sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|) \right) dm \geq \|\underline{q}\|_2 h(\underline{\alpha}) \int \log |T'| dm \geq \|\underline{q}\|_2 > 0.$$

In particular, the function in (2.3) tends to infinity as $\|\underline{q}\|_2 \rightarrow \infty$ and, by convexity of pressure, we can see that there is a unique minimum, attained at $\underline{q} = \underline{q}(\alpha)$, say. If μ_q is the Gibbs measure for

$$-h(\underline{\alpha}) \log |T'| + \sum_{i=1}^{2g} q_i(\alpha)(\bar{\psi}_i - \alpha_i \log |T'|),$$

then for each $1 \leq i \leq 2g$, the i th partial derivative at $\underline{q}(\alpha)$ is precisely

$$\int (\bar{\psi}_i - \alpha_i \log |T'|) d\mu_q,$$

which vanishes by assumption. Thus $\int \widehat{\Psi} d\mu_q / \int \log |T'(x)| d\mu_q = \underline{\alpha}$. Moreover,

$$\begin{aligned} & P \left(-h(\underline{\alpha}) \log |T'| + \sum_{i=1}^{2g} q_i(\bar{\psi}_i - \alpha_i \log |T'|) \right) \\ (2.4) \quad & = \underbrace{h(\mu_q) - h(\underline{\alpha}) \int \log |T'| d\mu_q}_{\leq 0} + \sum_{i=1}^{2g} q_i \underbrace{\int (\bar{\psi}_i - \alpha_i \log |T'|) d\mu_q}_{=0} \leq 0. \end{aligned}$$

Comparing (2.3) and (2.4) we see that

$$P\left(-\left(h(\underline{\alpha})+\sum_{i=1}^{2g}\alpha_i\right)\log|T'|+\sum_{i=1}^{2g}q_i\overline{\psi}_i\right)=0.$$

In particular, the result follows with $t(\underline{q})=(h(\underline{\alpha})+\sum_{i=1}^{2g}\alpha_i)$. \square

Finally, we have the following identities.

Lemma 8. *We have that: (1) $\alpha_j=\partial t/\partial q_j$; and (2) $\mu(\widehat{Y}_{\underline{\alpha}})=1$.*

Proof. For part (1) we know by the formula for the derivative of pressure [Ru] that

$$\frac{\partial}{\partial t}P(-t\log|T'|\circ\pi+q_1\widehat{\psi}_1+\cdots+q_{2g}\widehat{\psi}_{2g})|_{t=t(\underline{q})}=-\int\log|T'|\circ\pi d\mu$$

and

$$\frac{\partial}{\partial q_j}P(-t(\underline{q})\log|T'|\circ\pi+q_1\widehat{\psi}_1+\cdots+q_{2g}\widehat{\psi}_{2g})|_{q=q_j}=-\int\widehat{\psi}_jd\mu.$$

The result then follows from the implicit function theorem.

Part (2) follows from the Birkhoff ergodic theorem: For a.e. $(\mu)\ x$ we have that

$$\lim_{n\rightarrow+\infty}\frac{1}{n}\sum_{i=1}^n\log|T'\circ\pi(\sigma^ix)|=\int\log|T'|\circ\pi d\mu\text{ and }\lim_{n\rightarrow+\infty}\frac{1}{n}\sum_{i=1}^n\widehat{\psi}_j(\sigma^ix)=\int\widehat{\psi}_jd\mu,$$

for $j=1,\ldots,2g$. Using (2.2) we have that $\mu(\widehat{Y}_{\underline{\alpha}})=1$, as required. \square

3. POINTWISE DIMENSIONS

In this section we show that Lemma 4 follows from properties of pointwise dimension. This is a straightforward modification of the approach in [PW] for a single function. For convenience of notation, we denote $\nu=\pi^*\mu$ on I .

Definition. Given a point $\omega\in I$, we define the *upper pointwise dimension* and *lower pointwise dimension* to be

$$\overline{d}_\nu(\omega)=\limsup_{r\rightarrow 0}\frac{\log\nu(B(\omega,r))}{\log r}\text{ and }\underline{d}_\nu(\omega)=\liminf_{r\rightarrow 0}\frac{\log\nu(B(\omega,r))}{\log r},$$

respectively.

The next lemma gives bounds on these densities in terms of quantities we defined in the previous section.

Lemma 9.

- (i) *For a.e. $(\nu)\ \omega\in Y_{\underline{\alpha}}$ we can bound $\underline{d}_\nu(\omega)\geq t(\underline{q})+q_1\alpha_1+\cdots+q_{2g}\alpha_{2g}$.*
- (ii) *For all $\omega\in Y_{\underline{\alpha}}$ we can bound $\overline{d}_\nu(\omega)\leq t(\underline{q})+q_1\alpha_1+\cdots+q_{2g}\alpha_{2g}$.*

The proof of Lemma 9 is presented in the next section. Assuming this result, we can now complete the proof of Proposition 1.

Lemma 10. *For $\underline{\alpha}=(\alpha_1,\ldots,\alpha_{2g})\in\mathbb{R}^{2g}$ and $\underline{q}=(q_1,\ldots,q_{2g})\in\mathbb{R}^{2g}$ related as above, we can identify*

$$(3.1)\qquad\dim_H(Y_{\underline{\alpha}})=t(\underline{q})+q_1\alpha_1+\cdots+q_{2g}\alpha_{2g}.$$

Proof. It is a standard result that if there exists d such that $\underline{d}_\nu(\omega) \geq d$ for a.e. (ν) $\omega \in Y_\alpha$, then $\dim_H(Y) \geq d$ [Pe, Theorem 7.1]. Thus we deduce from Lemma 9 (i) that $\dim_H(Y) \geq t(q_1, \dots, q_{2g}) + q_1\alpha_1 + \dots + q_{2g}\alpha_{2g}$.

By a folklore theorem [Fa], if there exists d such that $\bar{d}_\nu(\omega) \leq d$ for every $\omega \in Y_\alpha$, this implies that $\dim_H(Y) \geq d$. Taking $d = t(\underline{q}) + q_1\alpha_1 + \dots + q_{2g}\alpha_{2g}$, by Lemma 9 (ii) we get that $\dim_H(Y_\alpha) \geq t(\underline{q}) + q_1\alpha_1 + \dots + q_{2g}\alpha_{2g}$. \square

The following lemma helps complete the proof of Lemma 4 (1).

Lemma 11. *The map $\alpha \mapsto \dim_H(Y_\alpha)$, $\alpha \in \text{int}(\mathcal{B})$, is analytic and strictly convex.*

Proof. The analyticity of $\dim_H(Y_\alpha)$ follows from the analyticity of $\underline{q}(\alpha)$, which in turn follows from its characterization (in the proof of Lemma 7), the analyticity of the pressure and the implicit function theorem.

The convexity of $\underline{q} \mapsto t(\underline{q})$ is easily checked to be strictly convex by showing that the second derivative is positive definite. More precisely, by analogy with the one-dimensional case [Pe, p. 212] we can compute

$$D^2t(\underline{q})(v_1, v_2) = \frac{D^2P(-\log|T'| - \langle Dt(v_1)\bar{\Psi} \rangle, -\log|T'| - \langle Dt(v_2)\bar{\Psi} \rangle)}{\int \log|T'|d\mu}.$$

The expression in (3.1) is the Legendre transform of $t(\underline{q})$ and thus, again, is strictly convex [Ar, p. 64]. \square

By convexity we see that $\dim_H(Y_\alpha) \leq 1$, with equality when $\alpha = 0$ [PW]. We can associate a Gibbs measure m with $m(X_\alpha) = 1$. Since Gibbs measures are fully supported, we can deduce that $Y_\alpha(T, S) \subset I$ is uncountable and dense, and thus $X_\alpha \subset S_1V$ is also dense and uncountable, as required.

4. PROOF OF LEMMA 9

We give the postponed proof of Lemma 9.

Proof of Lemma 9 (i). Fix $\epsilon > 0$. For each $w \in Y_\alpha$ we can choose $N(w)$ sufficiently large so that for $n \geq N(w)$ we have:

$$(4.1) \quad \alpha_j - \epsilon \leq \frac{\sum_{i=0}^{n-1} \psi_j(T^i\omega)}{\sum_{i=0}^{n-1} \log|T'(T^i\omega)|} \leq \alpha_j + \epsilon, \quad \forall 1 \leq j \leq 2g,$$

by Lemma 8 (2). For each $l \in \mathbb{N}$ let $Q_l = \{\omega : N(\omega) \leq l\}$ and observe that $Q_l \subset Q_{l+1}$ and $Y_\alpha = \bigcup_{l=1}^\infty Q_l$ (up to a set of zero measure). Choose l_0 sufficiently large that $\mu_q(Q_l) > 0$ whenever $l \geq l_0$.

For l_0 fixed, choose $l \geq l_0$. Given $0 < r < 1$, we can cover $\pi^{-1}(Q_l) \subset \Sigma_A^+$ by cylinders $C^{(i)}$, $i = 1, \dots, N$, of length $n(i)$ based at points $x_i \in \pi^{-1}(Q_l)$, say, and length $n = n(x, r)$ such that

$$(4.2) \quad |(T^{n(i)})'(\pi(x_i))| \leq r < |(T^{n(i)-1})'(\pi(x_i))|.$$

(This called a Moran cover [Pe].) Moreover, we can assume, without loss of generality, that the length of each cylinder is at least l_0 . Using successively (2.1), (4.1)

and (4.2) we can bound for any $\omega \in Q_l$:

$$\begin{aligned}
 \nu(B(\omega, r) \cap Q_l) &\leq \sum_{i=1}^N \mu(C^{(i)}) \\
 &\leq \sum_{i=1}^N C \left(|(T^{n(i)})'(\pi x_i)|^{-t(\underline{q})} e^{q_1 \hat{\psi}_1^{n(i)}(x_i) + \dots + q_{2g} \hat{\psi}_{2g}^{n(i)}(x_i)} \right) \\
 &\leq \sum_{i=1}^N C \left(|(T^{n(i)})'(\pi x_i)|^{-t(\underline{q}) - \sum_{j=1}^{2g} q_j (\alpha_j - \epsilon)} \right) \\
 &\leq C' r^{t(\underline{q}) + \sum_{j=1}^{2g} q_j (\alpha_j - \epsilon)},
 \end{aligned}
 \tag{4.3}$$

for some constant $C' > 0$. By the Borel density theorem we have that for ω in a full measure (i.e., essential) set of Q_l there exists $r_0(\omega) > 0$ such that

$$\nu(B(x, r)) \leq 2\nu(B(x, r) \cap Q_l), \text{ whenever } r < r_0(\omega).
 \tag{4.4}$$

In particular, comparing (4.3) and (4.4) shows that for any $l > l_0$ and a.e. $(\mu) \omega \in Q_l$ we have

$$d_\nu(\omega) = \liminf_{r \rightarrow 0} \frac{\log \nu(B(\omega, r))}{\log r} \geq t(\underline{q}) + \sum_{j=1}^{2g} q_j (\alpha_j - \epsilon).
 \tag{4.5}$$

Since l and $\epsilon > 0$ are arbitrary we deduce that (4.5) holds for a.e. $(\nu) \omega \in Y_\alpha$, as required. \square

Remark. Notice that the only place where we used that we have used $2g$ limits (rather than the usual single limit as in [PW]) is in (4.1). This only influences the definition $N(\omega)$ and, thus, that of Q_l , but does not change the usual proof in any other way.

Proof of Lemma 9 (ii). Fix $r > 0$. Let $\pi(x) = \omega \in I$ and choose $n = n(x, r)$ precisely so that

$$|(T^n)'(\omega)|^{-1} \leq r < |(T^{n-1})'(\omega)|^{-1}.
 \tag{4.6}$$

In particular, $\pi([x_0, \dots, x_{2g}]) \subset B(\omega, Dr)$, for some $D > 0$. Thus for all $\omega \in Y_\alpha$ we have by (2.1), (4.1) and (4.6),

$$\begin{aligned}
 \nu(B(\omega, Dr)) &\geq \mu([x_0, \dots, x_m]) \\
 &\geq \frac{1}{C} |(T^n)'(\pi x)|^{-t(\underline{q})} e^{q_1 \hat{\psi}_1^n(x) + \dots + q_{2g} \hat{\psi}_{2g}^n(x)} \\
 &\geq \frac{1}{C} |(T^n)'(\pi x)|^{-t(\underline{q}) - \sum_{j=1}^{2g} q_j (\alpha_j + \epsilon)} \\
 &\geq \frac{1}{C''} r^{t(\underline{q}) + \sum_{j=1}^{2g} q_j (\alpha_j + \epsilon)},
 \end{aligned}$$

for some constant $C'' > 0$. Thus

$$\bar{d}_\nu(x) = \limsup_{r \rightarrow 0} \frac{\log \nu_q(B(x, r))}{\log r} \leq t(\underline{q}) + \sum_{j=1}^{2g} q_j (\alpha_j + \epsilon).$$

Since $\epsilon > 0$ was arbitrary, we deduce that $\bar{d}_\nu(x) \leq t + \sum_{j=1}^{2g} q_j \alpha_j$. \square

Remark. A similar problem relates to the set $\mathcal{P} = \{\int z d\mu : \mu = T\text{-invariant}\} \subset \mathbb{C}$ of complex numbers which occur as the first Fourier coefficient of invariant measures for a C^2 expanding map $T : K \rightarrow K$ on the unit circle $K = \{z \in \mathbb{C} : |z| = 1\}$. The set

$$\mathcal{P} = \left\{ \int z d\mu(z) \in \mathbb{C} : \mu = T\text{-invariant probability} \right\}$$

has been extensively studied by Jenkinson [Je1], Bousch [Bo] and others, particularly in the case $T(z) = z^2$. For a typical point z on the unit circle the Birkhoff averages $\frac{1}{N} \sum_{n=1}^N T^n(z)$ converge to zero. We can denote $K_\omega = \{x \in X : \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N T^n(x) = \omega\}$, which has zero Liouville measure for $\omega \neq 0$. The analysis in this note shows that for $\underline{\alpha} \in \text{int}(\mathcal{P})$ we have that:

- (1) The Hausdorff dimension $\dim_H(X_\omega)$ satisfies $\dim_H(X_\omega) \leq 1$, with equality if and only if $\omega = 0$;
- (2) The Hausdorff dimension $\dim_H(X_{\underline{\alpha}})$ depends real analytically on $\underline{\alpha} \in \text{int}(\mathcal{P})$;
- (3) $X_\omega \subset I$ is dense, uncountable and carries a Gibbs measure.

For the particular case of the doubling map $T(z) = z^2$, this problem has been studied by Fan and Schmeling [FS] and the Birkhoff averages take the simple form $\frac{1}{N} \sum_{n=1}^N z^{2^k}$.

REFERENCES

- [Ar] V. Arnold, *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics, 60, Springer-Verlag, Berlin, 1978. MR **57**:14033b
- [Ba] V. Bangert, *Minimal measures and minimizing closed normal one-currents*, Geom. Funct. Anal. **9** (1999), 413–427. MR **2000m**:49058
- [BSS] L. Barreira, B. Saussol, and J. Schmeling, *Higher-dimensional multifractal analysis*, J. Math. Pures Appl. **81** (2002), 67–91.
- [Bo] T. Bousch, *Le poisson n'a pas d'arêtes*, Ann. Inst. Henri Poincaré Probab. Statist. **36** (2000), 489–508. MR **2001i**:37005
- [Bw1] R. Bowen, *Symbolic dynamics for hyperbolic flows*, Amer. J. Math. **95** (1973), 429–460. MR **49**:4041
- [Bw2] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics, Vol. 470, Springer-Verlag, Berlin, 1975. MR **56**:1364
- [BI] D. Burago and S. Ivanov, *Riemannian tori without conjugate points are flat*, Geom. Funct. Anal. **4** (1994), 259–269. MR **95h**:53049
- [Fa] K. Falconer, *Fractal geometry*, John Wiley, Chichester, 1990. MR **92j**:28008
- [FS] A. H. Fan and J. Schmeling, *On fast Birkhoff averaging*, Math. Proc. Cambridge Philos. Soc., to appear.
- [Fr] J. Franks, *Knots, links and symbolic dynamics.*, Annals of Math. **113** (1981), 529–552. MR **83h**:58074
- [Je1] O. Jenkinson, *Frequency locking on the boundary of the barycentre set*, Experiment. Math. **9** (2000), 309–317. MR **2001g**:37050
- [Je2] O. Jenkinson, *Rotation, Entropy, and Equilibrium States*, Trans. Amer. Math. Soc. **353** (2001), 3713–3739. MR **2002e**:37004
- [Ma1] D. Massart, *Ph.D. Thesis*, (Lyon).
- [Ma2] D. Massart, *Stable norms of surfaces: local structure of the unit ball of rational directions*, Geom. Funct. Anal. **7** (1997), 996–1010. MR **99b**:53061
- [MR] G. McShane and I. Rivin, *Simple curves on hyperbolic tori*, C. R. Acad. Sci. Paris Sér. I Math. **320** (1995), 1523–1528. MR **96g**:57018
- [Pe] Y. Pesin, *Dimension theory in dynamical systems.*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, 1997. MR **99b**:58003
- [PS] Y. Pesin and V. Sadovskaya, *Multifractal analysis of conformal Axiom A flows*, Comm. Math. Phys. **216** (2001), 277–312. MR **2002g**:37035

- [PW] Y. Pesin and H. Weiss, *The multifractal analysis of Gibbs measures: motivation, mathematical foundation, and examples*, *Chaos* **7** (1997), 89–106. MR **98e**:58130
- [Ra] M. Ratner, *Markov partitions for Anosov flows on n -dimensional manifolds*, *Israel J. Math.* **15** (1973), 92–114. MR **49**:4042
- [Sc] S. Schwartzman, *Asymptotic cycles*, *Ann. of Math.* **66** (1957), 270–284. MR **19**:568i
- [Ru] D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley Publ. Co., Reading, MA, 1978. MR **80g**:82017
- [Su] D. Sullivan, *Cycles for the dynamical study of foliated manifolds and complex manifolds*, *Invent. Math.* **36** (1976), 225–255. MR **55**:6440

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MANCHESTER, OXFORD ROAD, MANCHESTER, M13 9PL, ENGLAND

CONSTRUCTIONS PRESERVING HILBERT SPACE UNIFORM EMBEDDABILITY OF DISCRETE GROUPS

MARIUS DADARLAT AND ERIK GUENTNER

ABSTRACT. Uniform embeddability (in a Hilbert space), introduced by Gromov, is a geometric property of metric spaces. As applied to countable discrete groups, it has important consequences for the Novikov conjecture. Exactness, introduced and studied extensively by Kirchberg and Wassermann, is a functional analytic property of locally compact groups. Recently it has become apparent that, as properties of countable discrete groups, uniform embeddability and exactness are closely related. We further develop the parallel between these classes by proving that the class of uniformly embeddable groups shares a number of permanence properties with the class of exact groups. In particular, we prove that it is closed under direct and free products (with and without amalgam), inductive limits and certain extensions.

1. INTRODUCTION

Gromov introduced the notion of uniform embeddability of metric spaces, and suggested that finitely generated discrete groups that are uniformly embeddable in a Hilbert space, when viewed as metric spaces, might satisfy the Novikov conjecture [12], [10]. Yu proved that this is indeed the case [20], [18].

Kirchberg and Wassermann defined the notion of exactness of a locally compact group in terms of the behavior of its reduced crossed product functor. Subsequently, they developed the main properties of exact groups. In particular, they showed that in the case of a countable discrete group, exactness can be reformulated entirely in terms of the reduced C^* -algebra [14], that is, that exactness is a property of the harmonic analysis of the left regular representation of such a group.

The starting point of this work is the startling fact that for countable discrete groups, uniform embeddability (in a Hilbert space, a geometric property) and exactness (an analytic property) are closely related. The first indications of the relationship between uniform embeddability and exactness are found in the work of Guentner and Kaminker [13]; these preliminary steps were quickly expanded by Ozawa [16], Anantharaman-Delaroche [2], and others.

We are concerned with uniformly embeddable groups. Our main results are outlined in the following theorem (more precise statements follow in later sections),

Received by the editors July 22, 2002 and, in revised form, December 26, 2002.

2000 *Mathematics Subject Classification.* Primary 46L89, 20F65.

The first author was supported in part by an MSRI Research Professorship and NSF Grant DMS-9970223. The second author was supported in part by an MSRI Postdoctoral Fellowship and NSF Grant DMS-0071402.

which summarizes the basic permanence properties of the class of uniformly embeddable groups. Observe that the properties described are all shared by the class of countable discrete exact groups [15]. Indeed, in each case it is possible to give a unified account of the results for uniform embeddability and exactness; in some cases our methods provide alternate proofs of the known results concerning exactness.

Theorem. *The class of countable discrete groups that are uniformly embeddable in a Hilbert space is closed under subgroups and products, direct limits, free products with amalgam, and extensions by exact groups.* \square

The fact that subgroups and products of uniformly embeddable groups are again uniformly embeddable is elementary and quite well known; they are included in the statement for completeness.

The other properties are more difficult to establish. It is possible to construct a uniform embedding of a free product (without amalgam) directly from uniform embeddings of the factors [5]. On the other hand, the corresponding result for free products with amalgam is considerably more difficult, in view of the fact that the common subgroup of the amalgam can introduce considerable distortion into the product. Our proof is based on a suitable adaptation of an argument given by Tu in his work on Property A [19]; although we are not able to verify a number of assertions concerning the metric defined in Section 9 in Tu's paper, we are able to adapt his arguments to the present context. The proof we give works equally well for countable exact groups (see Proposition 6.8 and Theorem 6.9), and is unrelated to Dykema's original proof that the class of countable exact groups is closed under free products with amalgam [9], [8]. Again, in the case without amalgam a considerably simpler proof of this fact is now available [5].

The general problem of uniform embeddability of extensions is intriguing. Our proof that the class of uniformly embeddable groups is closed under extensions by exact groups is inspired by the argument of Anantharaman-Delaroche and Renault showing that the class of countable exact groups is closed under extensions [3]. It is unknown whether the class of uniformly embeddable groups is closed under general extensions; even the case of a central extension of \mathbb{Z} by a uniformly embeddable group remains open. At present, the behavior with respect to extensions provides the best possibility of distinguishing the classes of uniformly embeddable and exact groups.

We draw two immediate corollaries. Since they are peripheral to our study we will not establish notation or provide the relevant definitions; rather, we provide references.

Corollary. *The class of countable discrete groups that are uniformly embeddable in a Hilbert space is closed under the formation of HNN extensions.*

Proof. An HNN extension is built from free products with amalgam, direct limits, and a semi-direct product by \mathbb{Z} , which is exact [17], [4]. (See [6] for related results.) \square

Corollary. *The fundamental group of a graph of countable discrete groups is uniformly embeddable in a Hilbert space if and only if each of the groups is uniformly embeddable in a Hilbert space.*

Proof. Each constituent group is a subgroup of the fundamental group. Conversely, the fundamental group of a graph of groups is built from free products with amalgam, HNN extensions and direct limits [17], [4]. (See [1] for related remarks.) \square

2. BACKGROUND

Let X and Y be metric spaces, with metrics d_X and d_Y , respectively. A function $F : X \rightarrow Y$ is a *uniform embedding* if there exist non-decreasing functions $\rho_{\pm} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\lim_{t \rightarrow \infty} \rho_{\pm}(t) = \infty$ and such that

$$(1) \quad \rho_-(d_X(x, x')) \leq d_Y(F(x), F(x')) \leq \rho_+(d_X(x, x')), \quad \text{for all } x, x' \in X.$$

The space X is *uniformly embeddable* if there exists a uniform embedding F of X into a Hilbert space \mathcal{H} . Uniform embeddability in a real Hilbert space is equivalent to uniform embeddability in a complex Hilbert space; henceforth we shall deal only with real Hilbert spaces. Obviously, if X is countable we may assume that the Hilbert space is separable.

A metric space X is *locally finite* if for every $x \in X$ and $R > 0$ the metric ball with center x and radius R is finite. In this case the metric is called *proper*. A locally finite metric space is discrete (as a topological space in the metric topology). In the case of locally finite metric spaces there are a number of equivalent formulations of uniform embeddability [7], [13]; to these we add the following simple extension, which applies to *any* metric space and which will be our fundamental criterion for uniform embeddability.

Proposition 2.1. *Let X be a metric space. Then X is uniformly embeddable if and only if for every $R > 0$ and $\varepsilon > 0$ there exists a Hilbert space valued map $\xi : X \rightarrow \mathcal{H}$, $(\xi_x)_{x \in X}$, such that $\|\xi_x\| = 1$ for all $x \in X$, and such that*

- (i) $\sup\{\|\xi_x - \xi_{x'}\| : d(x, x') \leq R, x, x' \in X\} \leq \varepsilon,$
- (ii) $\lim_{S \rightarrow \infty} \sup\{|\langle \xi_x, \xi_{x'} \rangle| : d(x, x') \geq S, x, x' \in X\} = 0.$

These conditions may be replaced by

- (iii) $\sup\{|1 - \langle \xi_x, \xi_{x'} \rangle| : d(x, x') \leq R, x, x' \in X\} \leq \varepsilon,$
- (iv) $\lim_{S \rightarrow \infty} \inf\{\|\xi_x - \xi_{x'}\| : d(x, x') \geq S, x, x' \in X\} = 2,$

respectively.

Remark. We refer to (i) and (iii) collectively as the *convergence condition*; similarly, we refer to (ii) and (iv) collectively as the *support condition*.

Proof. The interchangeability of (i) \Leftrightarrow (iii) and (ii) \Leftrightarrow (iv) follows from the simple observation that for unit vectors $\xi, \eta \in \mathcal{H}$ we have $\|\xi - \eta\|^2 = 2 - 2\langle \xi, \eta \rangle$.

Assume that X is uniformly embeddable and let $F : X \rightarrow \mathcal{H}$ be a uniform embedding in a real Hilbert space \mathcal{H} . Let

$$\text{Exp}(\mathcal{H}) = \mathbb{R} \oplus \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H}) \oplus (\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}) \oplus \dots$$

and define $\text{Exp} : \mathcal{H} \rightarrow \text{Exp}(\mathcal{H})$ by

$$\text{Exp}(\zeta) = 1 \oplus \zeta \oplus \left(\frac{1}{\sqrt{2!}} \zeta \otimes \zeta \right) \oplus \left(\frac{1}{\sqrt{3!}} \zeta \otimes \zeta \otimes \zeta \right) \oplus \dots$$

Note that $\langle \text{Exp}(\zeta), \text{Exp}(\zeta') \rangle = e^{\langle \zeta, \zeta' \rangle}$, for all $\zeta, \zeta' \in \mathcal{H}$. For $t > 0$ define

$$\xi_x = e^{-t\|F(x)\|^2} \text{Exp}(\sqrt{2t} F(x)).$$

It is easily verified that $\langle \xi_x, \xi_{x'} \rangle = e^{-t\|F(x)-F(x')\|^2}$. Consequently, for all $x, x' \in X$ we have $\|\xi_x\| = 1$, and

(2)
$$e^{-t\rho_+(d(x,x'))^2} \leq \langle \xi_x, \xi_{x'} \rangle \leq e^{-t\rho_-(d(x,x'))^2}.$$

Letting $t = \varepsilon(1 + \rho_+(R)^2)^{-1}$, it is easy to verify the conditions (iii) and (ii) above.

Conversely, assume that X satisfies the conditions in the second part of the statement. There exist a sequence of maps $\eta_n : X \rightarrow \mathcal{H}_n$ and a sequence of numbers $S_0 = 0 < S_1 < S_2 < \dots$, increasing to infinity, such that for every $n \geq 1$ and every $x, x' \in X$,

- (i) $\|\eta_n(x)\| = 1$,
- (ii) $\|\eta_n(x) - \eta_n(x')\| \leq 1/n$, provided $d(x, x') \leq \sqrt{n}$,
- (iii) $\|\eta_n(x) - \eta_n(x')\| \geq 1$, provided $d(x, x') \geq S_n$.

Choose a base point $x_0 \in X$ and define $F : X \rightarrow \bigoplus_{n=1}^\infty \mathcal{H}_n$ by

$$F(x) = \frac{1}{2} (\eta_1(x) - \eta_1(x_0) \oplus \eta_2(x) - \eta_2(x_0) \oplus \dots).$$

It is not hard to verify that F is well defined and

$$\rho_-(d(x, x')) \leq \|F(x) - F(x')\| \leq d(x, x') + 1, \quad \text{for all } x, x' \in X,$$

where $\rho_- = \frac{1}{2} \sum_{n=1}^\infty \sqrt{n-1} \chi_{[S_{n-1}, S_n)}$, and the $\chi_{[S_{n-1}, S_n)}$ are the characteristic functions of the sets $[S_{n-1}, S_n)$.

Indeed, let $x, x' \in X$. If n is such that $\sqrt{n-1} \leq d(x, x') < \sqrt{n}$, we have

$$\begin{aligned} \|F(x) - F(x')\|^2 &= \frac{1}{4} \sum_{i \leq n-1} \|\eta_i(x) - \eta_i(x')\|^2 + \frac{1}{4} \sum_{i \geq n} \|\eta_i(x) - \eta_i(x')\|^2 \\ &\leq (n-1) + \frac{1}{4} \sum_{i \geq n} \frac{1}{i^2} \leq d(x, x')^2 + 1. \end{aligned}$$

Similarly, if n is such that $S_{n-1} \leq d(x, x') < S_n$, we have

$$\|F(x) - F(x')\|^2 \geq \frac{1}{4} \sum_{i \leq n-1} \|\eta_i(x) - \eta_i(x')\|^2 \geq \frac{n-1}{4} = \rho_-(d(x, x'))^2. \quad \square$$

Remark. Straightforward modifications of the above argument produce a uniform embedding F satisfying the sharper estimate $\|F(x) - F(x')\| \leq d(x, x') + \delta$, for an arbitrarily chosen $\delta > 0$; simply replace the $1/n$ in (ii) by $\delta/2^n$.

Two metrics d and d' on the set X are *coarsely equivalent* if for every $R > 0$ there exists an $S > 0$ such that the d -metric ball with center x and radius R is contained in the d' -metric ball with center x and radius S ; and conversely. Equivalently, two metrics on X are coarsely equivalent if the identity map $X \rightarrow X$ is a uniform embedding.

Proposition 2.2. *Let d and d' be coarsely equivalent metrics on X . Then X is uniformly embeddable with respect to d if and only if it is uniformly embeddable with respect to d' .*

Proof. If two maps are uniform embeddings, so is their composition. \square

Remark. Below we require only the fact that if $Y \rightarrow \mathcal{H}$ and $X \rightarrow Y$ are uniform embeddings, then the composite $X \rightarrow \mathcal{H}$ is a uniform embedding. In other words, we do not need to know that X and Y are coarsely equivalent to conclude the uniform embeddability of X from that of Y .

Let Γ be a countable discrete group. A *length function* on Γ is a nonnegative, real-valued function l satisfying, for all a and b in Γ ,

- (i) $l(ab) \leq l(a) + l(b)$,
- (ii) $l(a^{-1}) = l(a)$,
- (iii) $l(a) = 0$ if and only if $a = 1$.

A length function l is *proper* if for all $C > 0$ the subset $l^{-1}([0, C]) \subset \Gamma$ is finite. One can construct an integer-valued proper length function on Γ as follows. Let S be a symmetric set of generators of Γ . Let $l_0 : S \rightarrow \mathbb{N}$ be a proper function satisfying (ii) and (iii) above. Then

$$l(a) = \inf\{l_0(a_1) + \cdots + l_0(a_n) : a = a_1 \cdots a_n, a_i \in S\}$$

is a proper length function on Γ . Given a length function l , we define a metric d by $d(a, b) = l(a^{-1}b)$. A metric constructed in this way from a length function is *left-invariant* in the sense that $d(ca, cb) = d(a, b)$, for all a, b and $c \in \Gamma$. Conversely, every left-invariant metric arises in this way from a length function. A length function is proper if and only if the corresponding (left-invariant) metric has bounded geometry. Recall that a metric space X has *bounded geometry* if for every $R > 0$, there is a uniform bound on the number of elements in the balls of radius R in X .

We require the following well-known proposition (compare [19, Lemma 2.1]).

Proposition 2.3. *Let Γ be a countable discrete group and let d and d' be metrics on Γ associated to proper length functions l and l' , respectively. Then d and d' are coarsely equivalent.*

Proof. Since the metrics are left-invariant, it suffices to consider the containment, as in the definition, of balls centered at the identity element. By symmetry it suffices to show that for every $R > 0$ there exists an $S > 0$ such that, for all $a \in \Gamma$, if $l(a) < R$, then $l'(a) < S$. But, since l is proper, this is obvious. \square

As a consequence of the previous two propositions, uniform embeddability of a countable discrete group Γ is independent of the particular proper length function used to define its metric. Consequently, *we systematically omit reference to a specific length function or metric in the statements of all results* and say simply Γ is *uniformly embeddable* to mean that Γ is uniformly embeddable in a Hilbert space for some (equivalently all) left-invariant proper metric(s).

Finally, we draw two simple consequences. An action of a discrete group on a locally finite metric space X is *proper* if for every bounded subset $B \subset X$ the set $\{a \in \Gamma : a \cdot B \cap B \neq \emptyset\}$ is finite. Equivalently, for every $x \in X$ and $R \geq 0$ the set $\{a \in \Gamma : a \cdot x \in B_R(x)\}$ is finite. Observe that a free action of a discrete group on a locally finite metric space is proper.

Corollary 2.4. *Let Γ be a countable discrete group equipped with a left-invariant proper metric. Let X be a locally finite metric space equipped with a free isometric action of Γ . Then the inclusion $\Gamma \rightarrow X$ as an orbit is a uniform embedding.*

Proof. Let $x_0 \in X$. Since the action of Γ on X is by isometries, $l(a) = d_X(a \cdot x_0, x_0)$ defines a length function on Γ . Since the action is free and X is locally finite, l is a proper length function. Let d be the left-invariant metric associated to l .

According to the previous proposition, the original metric on Γ is coarsely equivalent to d , which is precisely what was to be proved. \square

Corollary 2.5. *Let X and Γ be as in the statement of the previous corollary. If X is uniformly embeddable, then so is Γ .* \square

Property A is a condition on metric spaces introduced by Yu [20]. We will work with the following characterization of Property A.

Proposition 2.6 ([19]). *A discrete metric space X with bounded geometry has Property A if and only if for every $R > 0$ and $\varepsilon > 0$ there exist an $S > 0$ and a Hilbert space valued function $\xi : X \rightarrow \mathcal{H}$ such that for all $x, x' \in X$ we have $\|\xi_x\| = 1$, and*

- (i) $d(x, x') \leq R \Rightarrow \|\xi_x - \xi_{x'}\| \leq \varepsilon$,
- (ii) $d(x, x') \geq S \Rightarrow \langle \xi_x, \xi_{x'} \rangle = 0$.

Equivalently, for every $R > 0$ and $\varepsilon > 0$ there exist an $S > 0$ and $\xi : X \rightarrow l^2(X)$ such that for all $x, x' \in X$ we have $\|\xi_x\| = 1$, (i) as above, and

- (iii) $\text{supp } \xi_x \subset B_S(x)$.

 \square

Remark. As in the case of uniform embeddability, we refer to (i) as the *convergence condition* and to (ii) and (iii) collectively as the *support condition*.

It is clear from the proposition that Property A is invariant under coarse equivalence. We refer the reader to [15] for an introduction to exact groups. Our interest in Proposition 2.6 is motivated by the following result, inspired by [13].

Theorem 2.7 ([16], and also [2]). *A countable discrete group Γ is exact if and only if Γ has Property A with respect to some (every) left-invariant proper metric.* \square

In analogy with Corollary 2.5 we have the following result, in which we do not assume that X itself has Property A.

Corollary 2.8. *Let Γ be a countable discrete group. Assume that Γ acts freely and isometrically on a locally finite metric space X satisfying the convergence and support conditions of Proposition 2.6. Then Γ is exact.*

Proof. Include $\Gamma \subset X$ as an orbit. The convergence and support conditions of Proposition 2.6 pass from X to the subspace Γ . Since Γ has bounded geometry, Proposition 2.6 applies. \square

3. LIMITS

We begin to establish the closure properties of the class of countable discrete uniformly embeddable groups. In this section we treat direct limits, our main result being the following proposition.

Proposition 3.1. *Let Γ be the limit of a directed system of countable discrete groups $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow \dots$ in which the maps $G_n \rightarrow G_{n+1}$ are injective. If each of the groups G_n is uniformly embeddable, then so is Γ .*

Proof. The proof is based on a method of extending Hilbert space valued functions from a subgroup to an ambient group. Specifically, let $\xi : G \rightarrow \mathcal{H}$ be a Hilbert space valued function on a subgroup G of a countable discrete group Γ . Choose and fix a family of coset representatives $x \in X \subset \Gamma$ for Γ/G ; having done so, we can express each element $a \in \Gamma$ uniquely as a product $x_a g_a$, where $x_a \in X$ and $g_a \in G$. The extension $\widehat{\xi}$ of ξ is defined by

$$\widehat{\xi} : \Gamma \rightarrow \mathcal{H} \otimes l^2(\Gamma/G) \cong \mathcal{H} \otimes l^2(X), \quad \widehat{\xi}_a = \xi_{g_a} \otimes \delta_{aG} \cong \xi_{g_a} \otimes \delta_{x_a}.$$

Now let Γ be a direct limit as in the statement of the theorem. Equip Γ with a proper length function l_Γ and associated metric d_Γ . Metrize each of the subgroups G_n as subspaces of Γ ; the metric and length function on G_n are simply the restriction of d_Γ and l_Γ . Observe that for $a, b \in \Gamma$ we have

$$a^{-1}b \in G_n \Leftrightarrow aG_n = bG_n \Leftrightarrow x_a = x_b \in \Gamma,$$

and that, in this case,

$$(3) \quad d_{G_n}(g_a, g_b) = d_\Gamma(g_a, g_b) = d_\Gamma(x_a g_a, x_b g_b) = d_\Gamma(a, b).$$

We show that Γ satisfies the convergence and support conditions of Proposition 2.1. Let $\varepsilon > 0$ and $R > 0$ be given. Obtain n such that if $a \in \Gamma$ has $l_\Gamma(a) \leq R$, then $a \in G_n$; this is possible because the length function l_Γ on Γ is assumed to be proper. According to the criterion for uniform embeddability, obtain a Hilbert space valued function $\xi : G_n \rightarrow \mathcal{H}$ such that for all $g, h \in G_n$ we have $\|\xi_g\| = 1$ and

- (i) if $d_{G_n}(g, h) \leq R$, then $\|\xi_g - \xi_h\| < \varepsilon$;
- (ii) $\forall \hat{\varepsilon} > 0 \exists S > 0$ such that if $d_{G_n}(g, h) \geq S$, then $|\langle \xi_g, \xi_h \rangle| < \hat{\varepsilon}$.

Let $\hat{\xi}$ be the extension of ξ to Γ defined above. Clearly, $\|\hat{\xi}_a\| = 1$ for all $a \in \Gamma$, and it remains to verify the conditions of Proposition 2.1. Let $a, b \in \Gamma$. For the convergence condition, assume that $l_\Gamma(a^{-1}b) = d_\Gamma(a, b) \leq R$. By our choice of n we have $a^{-1}b \in G_n$ and, according to (3), $d_{G_n}(g_a, g_b) = d_\Gamma(a, b) \leq R$. Therefore,

$$\|\hat{\xi}_a - \hat{\xi}_b\| = \|\xi_{g_a} \otimes \delta_{aG_n} - \xi_{g_b} \otimes \delta_{bG_n}\| = \|\xi_{g_a} - \xi_{g_b}\| < \varepsilon.$$

For the support condition let $\hat{\varepsilon} > 0$, obtain S as in (ii) above, and assume $d_\Gamma(a, b) \geq S$. According to

$$(4) \quad \langle \hat{\xi}_a, \hat{\xi}_b \rangle = \langle \xi_{g_a}, \xi_{g_b} \rangle \langle \delta_{aG_n}, \delta_{bG_n} \rangle = \begin{cases} \langle \xi_{g_a}, \xi_{g_b} \rangle, & \text{if } aG_n = bG_n, \\ 0, & \text{otherwise,} \end{cases}$$

we may assume $a^{-1}b \in G_n$. But then, according to (3) again, $d_{G_n}(g_a, g_b) = d_\Gamma(a, b)$ and $|\langle \hat{\xi}_a, \hat{\xi}_b \rangle| < \hat{\varepsilon}$. \square

Remark. The previous argument is easily adjusted to yield a new proof of the fact that a direct limit, as in the statement of the proposition, of exact groups is again exact [15]. Indeed, under the assumption of exactness, employing Proposition 2.6 instead of Proposition 2.1, we replace (ii) by

$$\exists S > 0 \text{ such that } d_{G_n}(g, h) \geq S \Rightarrow \langle \xi_g, \xi_h \rangle = 0.$$

Using (4) and the surrounding discussion, we conclude that this property is shared by $\hat{\xi}$.

4. EXTENSIONS

Let $1 \rightarrow H \rightarrow \Gamma \rightarrow G \rightarrow 1$ be an extension of countable discrete groups. We study uniform embeddability of Γ under various hypotheses on H and G . Our primary result, of which our other results are consequences, is the following theorem.

Theorem 4.1. *Let Γ be an extension of H by G as above. If H is uniformly embeddable and G is exact, then Γ is uniformly embeddable.*

As corollaries we mention the following two results about semi-direct products.

Corollary 4.2. *Let G and H be countable discrete groups. Let $\alpha : G \rightarrow \text{Aut}(H)$ be an action of G on H . If both H and G are uniformly embeddable and $\alpha(G) \subset \text{Aut}(H)$ is exact, then the semi-direct product $\Gamma = H \rtimes G$ is uniformly embeddable.*

Proof. The semi-direct product Γ is the set of pairs $(s, x) \in H \times G$, with product $(s, x)(t, y) = (s\alpha_x(t), xy)$. The assignment

$$(s, x) \mapsto ((s, \alpha_x), x) : \Gamma \rightarrow (H \rtimes \alpha(G)) \times G$$

defines an injective homomorphism.

By the theorem the semi-direct product $H \rtimes \alpha(G)$ is uniformly embeddable, as is $(H \rtimes \alpha(G)) \times G$. In particular, Γ is a subgroup of the uniformly embeddable group and is therefore uniformly embeddable. \square

Corollary 4.3. *Let $\Gamma = \mathbb{Z}^n \rtimes G$ be a semi-direct product. If G is uniformly embeddable, then Γ is uniformly embeddable.*

Proof. Apply the previous corollary, using the fact that $GL_n(\mathbb{Z})$ is exact [15]. \square

Remark. Central extensions are more difficult to analyze than semi-direct products. Our theorem applies to central extensions in which the quotient is exact. Gersten has shown that if a central extension of \mathbb{Z} is described by a bounded cocycle, then Γ is quasi-isometric to the product $\mathbb{Z} \times G$ [11]. Consequently, if G is uniformly embeddable, so is Γ . Beyond these two results, little is known.

As remarked earlier, the property of uniform embeddability of a countable discrete group is independent of the proper length function. Consequently, we are free to choose these for our groups G , H and Γ in a convenient manner, which we do as follows. Let l_Γ be a proper length function on Γ . Define length functions on H and G according to

$$(5) \quad l_H(s) = l_\Gamma(s), \quad \text{for all } s \in H,$$

$$(6) \quad l_G(x) = \min\{l_\Gamma(a) : a \in \Gamma \text{ and } \dot{a} = x\}, \quad \text{for all } x \in G,$$

where we have introduced the notation $a \mapsto \dot{a}$ for the quotient map $\Gamma \rightarrow G$. It is easily verified that the minimum in the definition of l_G is attained, and that l_G is a proper length function on G ; that l_H is a proper length function is immediate. Denote the associated left-invariant metrics by d_Γ , d_G and d_H , respectively. Observe that the inclusion $H \hookrightarrow \Gamma$ is an isometry, and the quotient map $\Gamma \rightarrow G$ is contractive. Finally, choose a set-theoretic section σ of the quotient map $\Gamma \rightarrow G$ with the property that

$$(7) \quad l_\Gamma(\sigma(x)) = l_G(x), \quad \text{for all } x \in G,$$

and define $\eta : \Gamma \times G \rightarrow H$ by

$$(8) \quad \eta(a, x) = \sigma(x)^{-1}a\sigma(\dot{a}^{-1}x), \quad \text{for all } a \in \Gamma, x \in G.$$

Lemma 4.4. *Let $a, b \in \Gamma$, $x \in G$. We have*

$$(9) \quad d_\Gamma(a, b) \leq d_G(x, \dot{a}) + d_G(x, \dot{b}) + d_H(\eta(a, x), \eta(b, x)),$$

$$(10) \quad d_H(\eta(a, x), \eta(b, x)) \leq d_G(x, \dot{a}) + d_G(x, \dot{b}) + d_\Gamma(a, b).$$

Proof. Let a, b and $x \in G$ be as in the statement. From the definition (8) of η we obtain

$$\eta(a, x)^{-1}\eta(b, x) = \sigma(\dot{a}^{-1}x)^{-1}a^{-1}b\sigma(\dot{b}^{-1}x).$$

The desired inequalities follow easily from this equality, together with (5), (7) and the subadditivity of length functions. \square

Proof of Theorem 4.1. We prove that Γ satisfies the conditions of Proposition 2.1. Let $\varepsilon > 0$ and $R > 0$ be given. We show that there exists an $f : \Gamma \rightarrow l^2(G, \mathcal{H})$ such that $\|f(a)\| = 1$ for all $a \in \Gamma$, and that f satisfies the convergence and support properties:

- (i) if $d_\Gamma(a, b) \leq R$, then $|1 - \langle f(a), f(b) \rangle| < \varepsilon$;
- (ii) $\forall \hat{\varepsilon} > 0 \exists S > 0$ such that if $d_\Gamma(a, b) \geq S$, then $|\langle f(a), f(b) \rangle| < \hat{\varepsilon}$.

Adapting an argument of Anantharaman-Delaroche and Renault [3], our strategy is to use a g satisfying the conditions of Proposition 2.6 to average an h satisfying those of Proposition 2.1 to produce f . Since G is exact, there exist, according to Proposition 2.6, a $g : G \rightarrow l^2(G)$ and an $S_G > 0$ such that $\|g(x)\| = 1$ for all $x \in G$ and such that

- (i_g) $|1 - \langle g(x), g(y) \rangle| < \varepsilon/2$, provided $d_G(x, y) \leq R$;
- (ii_g) $\text{supp } g(x) \subset B_{S_G}(x)$, for all $x \in G$.

We shall without comment view g as a function on $G \times G$ whenever convenient. Since H is uniformly embeddable, there exists, according to Proposition 2.1, an $h : H \rightarrow \mathcal{H}$ such that $\|h(s)\| = 1$ for all $s \in H$ and such that

- (i_h) $|1 - \langle h(s), h(t) \rangle| < \varepsilon/2$, provided $d_H(s, t) \leq 2S_G + R$;
- (ii_h) $\forall \hat{\varepsilon} > 0 \exists S_H > 0$ such that if $d_H(s, t) \geq S_H$, then $|\langle h(s), h(t) \rangle| < \hat{\varepsilon}$.

Having chosen g and h , define $f : \Gamma \rightarrow l^2(G, \mathcal{H})$ by

$$f(a)(x) = g(\dot{a}, x)h(\eta(a, x)), \quad \text{for all } a \in \Gamma, x \in G.$$

Note that $f(a) \in l^2(G, \mathcal{H})$. It is elementary to verify that $\|f(a)\| = 1$, for all $a \in \Gamma$. We verify the remaining properties.

For the convergence property, let $a, b \in \Gamma$ with $d_\Gamma(a, b) \leq R$. Consider

$$(11) \quad |1 - \langle f(a), f(b) \rangle| \leq \left\{ \left| \sum_{x \in G} \{1 - \langle h(\eta(a, x)), h(\eta(b, x)) \rangle\} g(\dot{a}, x)g(\dot{b}, x) \right| + |1 - \langle g(\dot{a}), g(\dot{b}) \rangle| \right\}.$$

Since the map $\Gamma \rightarrow G$ is contractive we have $d_G(\dot{a}, \dot{b}) \leq R$, so that by (i_g) the second term in (11) is bounded by $\varepsilon/2$. Observe that, according to (ii_g), the sum in the first term is over $x \in B_{S_G}(\dot{a}) \cap B_{S_G}(\dot{b})$. Recalling that $\|g(\dot{a})\| = \|g(\dot{b})\| = 1$, we therefore bound the first term by

$$\sup\{|1 - \langle h(\eta(a, x)), h(\eta(b, x)) \rangle| : x \in B_{S_G}(\dot{a}) \cap B_{S_G}(\dot{b})\}.$$

From (10) we see that for $x \in B_{S_G}(\dot{a}) \cap B_{S_G}(\dot{b})$ we have $d_H(\eta(a, x), \eta(b, x)) \leq 2S_G + R$, so that by (i_h) this supremum, and consequently the first term in (11), is bounded by $\varepsilon/2$.

For the support property, let $\hat{\varepsilon} > 0$ be given and obtain S_H as in (ii_h). Let $a, b \in \Gamma$ be such that $d_\Gamma(a, b) \geq 2S_G + S_H$. Then,

$$\begin{aligned} |\langle f(a), f(b) \rangle| &= \left| \sum_{x \in G} g(\dot{a}, x) g(\dot{b}, x) \langle h(\eta(a, x)), h(\eta(b, x)) \rangle \right| \\ &\leq \sum_x |g(\dot{a}, x) g(\dot{b}, x)| |\langle h(\eta(a, x)), h(\eta(b, x)) \rangle| \\ &\leq \sup\{ |\langle h(\eta(a, x)), h(\eta(b, x)) \rangle| : x \in B_{S_G}(\dot{a}) \cap B_{S_G}(\dot{b}) \}, \end{aligned}$$

where again we use the fact that $\|g(\dot{a})\| = \|g(\dot{b})\| = 1$ to obtain the second inequality, and note that by (ii_g) the sums are in fact over $x \in B_{S_G}(\dot{a}) \cap B_{S_G}(\dot{b})$. From (9) we see that for such x we have $d_H(\eta(a, x), \eta(b, x)) \geq d_\Gamma(a, b) - d_G(x, \dot{a}) - d_G(x, \dot{b}) \geq S_H$, so that by (ii_h) the supremum is indeed bounded by $\hat{\varepsilon}$. \square

5. FREE PRODUCTS

The main result of this section is the following theorem.

Theorem 5.1. *Let A and B be countable discrete groups and let C be a common subgroup. If both A and B are uniformly embeddable, then the amalgamated free product $A *_C B$ is uniformly embeddable.*

Our strategy for proving the theorem is to construct a locally finite metric space X that is uniformly embeddable, and on which Γ acts freely by isometries. The construction of X is based on the notion of a tree of metric spaces, which we now recall.

A *tree* T consists of two sets, a set V of vertices and a set E of edges, together with two *endpoint maps* $E \rightarrow V$ associating to each edge its endpoints. Every two vertices are connected by a unique geodesic edge path, that is, a path without backtracking.

A *tree of spaces* \mathcal{X} (with base the tree T) consists of a family of metric spaces $\{X_v, X_e\}$ indexed by the vertices $v \in V$ and edges $e \in E$ of T together with maps $\sigma_{e,v} : X_e \rightarrow X_v$ whenever v is an endpoint of e . The $\sigma_{e,v}$ are the *structural maps* of \mathcal{X} . We will assume, although this is not strictly necessary, that the metrics on the vertex and edge spaces are integer-valued.

The *total space* X of the tree of spaces \mathcal{X} is the metric space defined as follows. The underlying set of X is the disjoint union of the vertex spaces X_v ; the metric on X is the metric envelope d of the partial metric \hat{d} (see the appendix) defined by

$$\hat{d}(x, y) = \begin{cases} d(x, y), & \text{if } \exists v \in V \text{ such that } x, y \in X_v, \\ 1, & \text{if } \exists e \in E \text{ and } z \in X_e \text{ such that } x = \sigma_{e,v}(z), y = \sigma_{e,w}(z), \end{cases}$$

for all (x, y) in the domain

$$\mathcal{D} = \{(x, y) : x, y \in X_v, v \in V\} \cup \{(\sigma_{e,v}(z), \sigma_{e,w}(z)) : v \neq w, z \in X_e, e \in E\}.$$

Observe that \mathcal{D} is ample (see the appendix); this follows from our assumption that the underlying tree is connected, and that \hat{d} is defined for all pairs (x, y) where x and y are in the same vertex space.

We call (x, y) an *adjacency* if there exist an edge e , with endpoints v and w , and an element $z \in X_e$, such that $\sigma_{e,v}(z) = x$ and $\sigma_{e,w}(z) = y$. Using this terminology,

the partial metric can be described as being the given metric on each vertex space and 1 on adjacencies.

Example 5.2. Consider the case in which the vertex spaces are metric graphs (or rather the set of vertices of a graph, equipped with the graph metric) and the edge spaces are singletons. The total space X is itself a metric graph. Indeed, it is the disjoint union of the graphs, together with additional edges coming from the underlying tree. Precisely, the edges of X are, first, the edges in the individual X_v , and second, the edges e of the underlying tree; the endpoints of such an edge e are the images of the maps $X_e \rightarrow X_v$ to the endpoints v of e in T .

Remark. The generality of the definition above is mandated by the fact that when considering amalgamated free products, we will encounter vertex spaces that are not metric graphs.

Theorem 5.3. *Let \mathcal{X} be a tree of metric spaces in which the structural maps are isometries. If the vertex spaces X_v are uniformly embeddable in a Hilbert space (with common distortion bound), then the total space X is uniformly embeddable in a Hilbert space.*

Remark. Quite explicitly, the hypothesis is that there exist maps ρ_{\pm} as in (1) such that for every vertex $v \in V$ there exists $F_v : X_v \rightarrow \mathcal{H}$ satisfying

$$\rho_-(d_{X_v}(x, x')) \leq \|F_v(x) - F_v(x')\| \leq \rho_+(d_{X_v}(x, x')), \quad \text{for all } x, x' \in X_v.$$

In other words, the same distortion bounds may be used for every vertex space.

Remark. If there exist only finitely many distinct isometry types of vertex spaces (as will be the case for the amalgamated free product), this simply means that every vertex space is uniformly embeddable.

Remark. When we prove the theorem we will use the fact that the existence of a common distortion bound on the uniform embeddings of the vertex spaces implies that the estimates in the fundamental criterion for embeddability are uniform. This follows from the proof of Proposition 2.1, specifically from the estimate (2).

Assuming Theorem 5.3, we prepare for the proof Theorem 5.1 by associating a tree of metric spaces \mathcal{X}_Γ to the amalgamated free product $\Gamma = A *_C B$. The tree T is the Bass-Serre tree of Γ [17], [4]. Precisely, the vertex and edge sets of T are given by

$$\begin{aligned} V &= \Gamma/A \cup \Gamma/B, \\ E &= \Gamma/C, \end{aligned}$$

respectively; the endpoint maps are the quotient maps $\Gamma/C \rightarrow \Gamma/A$ and $\Gamma/C \rightarrow \Gamma/B$. In other words, the endpoints of an edge (a C -coset) are the vertices (one A -coset and one B -coset) that contain it.

We associate a metric space to each vertex and edge of T as follows. Equip Γ with an integer-valued proper length function and associated metric. Let $v \in V$ be a vertex and assume that $v \in \Gamma/A$. In particular, v is an A -coset in Γ ; denote by X_v this coset itself, metrized as a subspace of Γ . Proceed similarly for vertices $v \in \Gamma/B$. Let $e \in E$ be an edge. In particular, e is a C -coset in Γ ; denote by X_e this coset itself, again metrized as a subspace of Γ .

The structural maps are defined as follows. Let the vertex v be an endpoint of the edge e . Inclusion of cosets (subsets of Γ) provides the structural map $\sigma_{e,v} : X_e \rightarrow X_v$.

Remark. The metric space X_v is *not* (isometric to) the graph metric space of the Cayley graph of A or B , or even of the restriction of the Cayley graph of Γ to the subset A or B . Similar remarks apply to the edge space X_e . Indeed, *it is important that we metrize X_v and X_e as subspaces of Γ , and not via their identification with A or B and C using the given metrics on these groups.*

Let X_Γ be the total space of \mathcal{X}_Γ . The group Γ acts on X_Γ by left multiplication. Precisely, for $x \in X_v$ and $a \in \Gamma$ we define $a \cdot x \in X_{a \cdot v}$, using ordinary multiplication in Γ , by virtue of the fact that $x \in v \subset \Gamma$. This action preserves adjacencies and, since in addition the metric on Γ is left-invariant, it preserves the partial metric. According to Proposition 7.2 of the appendix, Γ acts by isometries on the total space X_Γ .

Proposition 5.4. *Let A and B be countable discrete groups and let C be a common subgroup. Let $\Gamma = A *_C B$ be the amalgamated free product. Let \mathcal{X}_Γ be the tree of metric spaces associated to Γ and let X_Γ be its total space. We have:*

- (i) *The structural maps of \mathcal{X}_Γ are isometries.*
- (ii) *Every vertex space of \mathcal{X}_Γ is isometric to one of A or B (which are metrized with the subspace metric via the inclusions $A, B \subset \Gamma$).*
- (iii) *The action of Γ by isometries on X_Γ is free.*
- (iv) *X_Γ is locally finite.*

Proof. All of the assertions but (iv) follow from the previous discussion. The condition (iv), while apparent, will be derived formally in Proposition 5.6. \square

Proof of Theorem 5.1. According to the previous proposition, the tree of metric spaces \mathcal{X}_Γ associated to the amalgamated free product $\Gamma = A *_C B$ satisfies the hypothesis of Theorem 5.3, according to which the total space X_Γ is uniformly embeddable. Again according to the previous proposition, Γ acts freely by isometries on X_Γ . Hence, by Corollary 2.5, Γ is uniformly embeddable. \square

We have reduced our main theorem concerning amalgamated free products, Theorem 5.1, to a theorem concerning trees of metric spaces, Theorem 5.3. We now turn attention to the proof of that theorem. We suitably adapt the method employed by Tu in his study of Property A for discrete metric spaces [19].

For a tree of metric spaces \mathcal{X} the inclusions $X_v \rightarrow X$ of vertex spaces into the total space are, as a general rule, *not* isometric; the presence of “shortcuts” in neighboring X_w may cause the distance in X between two vertices $x, y \in X_v$ to be considerably smaller than the distance between them in X_v itself. Nevertheless, for our \mathcal{X}_Γ these inclusions are isometries, a fact that may be traced back to the manner in which the vertex and edge spaces of \mathcal{X}_Γ are metrized as subspaces of Γ , which circumvents any distortion that may have otherwise been introduced by the amalgamating subgroup. The following proposition has no analog in Tu’s work [19]; nevertheless, we require it in order to complete our arguments.

Proposition 5.5. *Let $\mathcal{X} = \{X_v, X_e\}$ be a tree of metric spaces in which the structural maps $X_e \rightarrow X_v$ are isometries. Let X be the total space of \mathcal{X} . Then*

- (i) *the inclusions $X_v \rightarrow X$ are isometries, and*

(ii) if (x, y) is an adjacency, then $d(x, y) = 1$.

Further, for all $x \in X_v$, $y \in X_w$,

$$(12) \quad d(x, y) = d_T(v, w) + \inf\{d(x_0, x_1) + d(x_2, x_3) + \cdots + d(x_{p-1}, x_p)\},$$

where d_T is the distance on the tree T and the infimum is taken over all sequences x_0, \dots, x_p , where $p = 2d_T(v, w) + 1$, and

(i) $x = x_0$, $y = x_p$,

(ii) (x_{2k-1}, x_{2k}) is an adjacency for $k = 1, \dots, d_T(v, w)$,

(iii) $x_{2k}, x_{2k+1} \in X_{v_k}$, for $k = 0, \dots, d_T(v, w)$, and

$v = v_0, \dots, v_{d_T(v, w)} = w$ are the vertices along the unique geodesic path in T from v to w .

Proof. Let $v, w \in V$ and let $x \in X_v$ and $y \in X_w$. A reduced path from x to y is a sequence of elements $x = x_0, x_1, \dots, x_{2n}, x_{2n+1}$ of X for which there exist vertices $v_0, \dots, v_n \in V$ such that

(i) $v_j \neq v_{j+1}$, for $j = 1, \dots, n$, and

(ii) $x_{2k}, x_{2k+1} \in X_{v_k}$, for $k = 0, 1, \dots, n$.

Observe that by appropriately inserting and deleting x 's we may alter a path, without increasing its length, so as to obtain a reduced path (use the triangle inequality for the metrics on the individual X_v). Consequently, when computing distances in X it suffices to consider reduced paths.

Let $x = x_0, \dots, x_{2n+1} = y$ be a reduced path from x to y . According to the definitions of a reduced path and of the domain \mathcal{D} of the partial metric we obtain sequences of

(i) vertices $v = v_0, \dots, v_n = w$ in V ,

(ii) edges e_1, \dots, e_n in E , and

(iii) elements z_1, \dots, z_n of the edge spaces X_{e_1}, \dots, X_{e_n} ,

satisfying

(i) $x_{2k}, x_{2k+1} \in X_{v_k}$, for $k = 0, \dots, n$,

(ii) the endpoints of e_k are v_{k-1} and v_k , for $k = 1, \dots, n$, and

(iii) $\sigma_{e_k, v_{k-1}}(z_k) = x_{2k-1}$ and $\sigma_{e_k, v_k}(z_k) = x_{2k}$, for $k = 1, \dots, n$.

(The sequences are uniquely determined by these conditions.) The given reduced path $x = x_0, \dots, x_p = y$ in X lies over the edge path e_1, \dots, e_n in T .

We show that in the definition of d it suffices to consider reduced paths lying over the unique geodesic edge path in T from v to w ; the assertions of the proposition follow easily from this fact. Let $x = x_0, \dots, x_{2n+1} = y$ lie over the non-geodesic path e_1, \dots, e_n in T . We show that by successive elimination of certain x_i we obtain a shorter path lying over the geodesic path in T . Indeed, there exists $i \in 1, \dots, n-1$ such that e_i and e_{i+1} have the same endpoints, that is, such that $v_{i-1} = v_{i+1}$. We have

$$\left. \begin{matrix} x_{2i-2} \\ x_{2i-1} \end{matrix} \right\} \in X_{v_{i-1}}, \quad \left. \begin{matrix} x_{2i} \\ x_{2i+1} \end{matrix} \right\} \in X_{v_i}, \quad \left. \begin{matrix} x_{2i+2} \\ x_{2i+3} \end{matrix} \right\} \in X_{v_{i+1}} = X_{v_{i-1}}.$$

We eliminate $x_{2i-1}, x_{2i}, x_{2i+1}, x_{2i+2}$ from the given path and claim that the resulting path

(i) is shorter than x_0, \dots, x_{2n+1} , and

(ii) lies over an edge path with fewer backtracks than e_1, \dots, e_n .

Of these, (ii) is obvious; the new path lies over the edge path $e_1, \dots, e_{i-1}, e_{i+2}, \dots, e_n$ obtained by eliminating e_i and e_{i+1} . Further, (i) follows from the fact that the structural maps are isometries. In particular, we have

$$d(x_{2i-1}, x_{2i+2}) = d(z_i, z_{i+1}) = d(x_{2i}, x_{2i+1}),$$

from which follows

$$\begin{aligned} d(x_{2i-2}, x_{2i+3}) &\leq d(x_{2i-2}, x_{2i-1}) + d(x_{2i-1}, x_{2i+2}) + d(x_{2i+2}, x_{2i+3}) \\ &= d(x_{2i-2}, x_{2i-1}) + d(x_{2i}, x_{2i+1}) + d(x_{2i+2}, x_{2i+3}). \end{aligned}$$

□

Proposition 5.6. *Let X be the total space of a tree of metric spaces \mathcal{X} in which the structural maps are isometries. Assume that each vertex space X_v is locally finite and there is a uniform bound for the number of adjacencies of each point in X . Then X is locally finite.*

Proof. In view of formula (12), a finite radius ball in X can intersect only finitely many vertex spaces. □

6. PROOF OF THEOREM 5.3

The main result of this section is Proposition 6.8, which implies Theorem 5.3 and at the same time reproves the exactness of free products with amalgam of exact groups.

We may enlarge the tree of metric spaces (if necessary) so that the underlying tree T will contain an infinite geodesic ω starting at some basepoint. For every $v \in V$ let $\alpha(v) \in V$ be such that the edge $[v, \alpha(v)]$ points towards ω . For each $v \in V$ let $Y_v = \sigma_{e,v}(X_e) \subset X_v$ and $f_v = \sigma_{e,\alpha(v)} \circ \sigma_{e,v}^{-1} : Y_v \rightarrow X_{\alpha(v)}$ where $e = [v, \alpha(v)]$. It follows immediately from Proposition 5.5 that each f_v is isometric.

Using the same proposition, and the notation just introduced, we rewrite the distance formula (12) as follows. Let $x_0 \in X_v$ and $x'_0 \in X_{v'}$. There exists a unique pair of nonnegative integers k, ℓ such that $\alpha^k(v) = \alpha^\ell(v')$ and $d_T(v, v') = k + \ell$, where again d_T denotes the distance in the tree T . By symmetry we may assume that $k \geq \ell$. If $k \geq 1$ and $\ell \geq 1$, then

$$(13) \qquad d(x_0, x'_0) = k + \ell + \inf \left\{ \begin{aligned} & d(x_0, y_0) + \sum_{i=0}^{k-2} d(f_{\alpha^i(v)}(y_i), y_{i+1}) \\ & \qquad \qquad \qquad + d(f_{\alpha^{k-1}(v)}(y_{k-1}), f_{\alpha^{\ell-1}(v')}(y'_{\ell-1})) \\ & \qquad \qquad \qquad + \sum_{j=0}^{\ell-2} d(f_{\alpha^j(v')}(y'_j), y'_{j+1}) + d(y'_0, x'_0) \end{aligned} \right\},$$

subject to the constraints that $y_i \in Y_{\alpha^i(v)}$ and $y'_j \in Y_{\alpha^j(v')}$. If $k \geq 1$ and $\ell = 0$, then

$$(14) \qquad d(x_0, x'_0) = k + \inf \left\{ d(x_0, y_0) + \sum_{i=0}^{k-2} d(f_{\alpha^i(v)}(y_i), y_{i+1}) + d(f_{\alpha^{k-1}(v)}(y_{k-1}), x'_0) \right\},$$

subject to similar constraints.

Remark. Formulas (13) and (14) are the same as Tu’s formulas [19, Section 9].

An n -chain is a sequence $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ with $x_k \in X_{\alpha^k(v)}$ such that for each $0 \leq k \leq n-2$ there exists $\bar{x}_k \in Y_{\alpha^k(v)}$ satisfying $d(x_k, \bar{x}_k) < d(x_k, Y_{\alpha^k(v)}) + 1$ and $f_{\alpha^k(v)}(\bar{x}_k) = x_{k+1}$. If $x_0 \in X_v$, the n -chain starts in X_v . Note that for any $n \geq 1$, $v \in V$ and $x_0 \in X_v$ there exists an n -chain whose initial element is x_0 .

Lemma 6.1. *Let $x_0 \in X_v$ and $x'_0 \in X_{v'}$ with $d(x_0, x'_0) < R$, and let k and ℓ be as in (13) and (14). Then there exist chains (x_0, x_1, \dots, x_k) , $(x'_0, x'_1, \dots, x'_\ell)$ such that*

$$\max \left\{ \left(\sup_{0 \leq i \leq k-1} d(x_i, x_{i+1}) \right), \left(\sup_{0 \leq j \leq \ell-1} d(x'_j, x'_{j+1}) \right), d(x_k, x'_\ell) \right\} < 2^R R.$$

Remark. In the proof of the lemma, and at a number of subsequent points, we require the fact that if $x \in X_v$ and $y \in Y_v$, then

$$d(x, f_v(y)) = d(x, y) + 1.$$

This follows from Theorem 5.5. Indeed, since $(y, f_v(y))$ is an adjacency, $d(x, f_v(y)) \leq d(x, y) + 1$. For the reverse inequality, let $\varepsilon > 0$ and obtain $x' \in Y_v$ such that $d(x, f_v(y)) + \varepsilon \geq 1 + d(x, x') + d(f_v(x'), f_v(y))$. Since f_v is an isometry, we get $d(x, x') + d(f_v(x'), f_v(y)) \geq d(x, y)$, and we are done.

Proof. If $v = v'$ there is nothing to prove; so we may assume that $v \neq v'$. By symmetry we may also assume that $k \geq \ell$; hence $k \geq 1$ since $v \neq v'$.

Case $\ell = 0$: We need to prove that there exists a chain (x_0, x_1, \dots, x_k) such that

$$\max \left\{ \left(\sup_{0 \leq i \leq k-1} d(x_i, x_{i+1}) \right), d(x_k, x'_0) \right\} < 2^k R \leq 2^R R.$$

Since $d(x_0, x'_0) < R$, by (14) there is $y_0 \in Y_v$ such that $k + d(x_0, y_0) < R$. The sequence x_1, \dots, x_k is constructed inductively. Let $\bar{x}_0 \in Y_v$ be such that $d(x_0, \bar{x}_0) \leq d(x_0, y_0)$ and $d(x_0, \bar{x}_0) < d(x_0, Y_v) + 1$, and define $x_1 = f_v(\bar{x}_0)$. Then

$$d(x_0, x_1) = d(x_0, \bar{x}_0) + 1 \leq d(x_0, y_0) + 1 < R.$$

Thus $d(x_0, x_1) < R$ and $d(x_1, x'_0) \leq d(x_0, x'_0) + d(x_0, x_1) < 2R$. Repeating the same argument for the pair of points x_1, x'_0 (if $k \geq 2$), we find $x_2 \in X_{\alpha^2(v)}$ with (x_0, x_1, x_2) being a chain, $d(x_1, x_2) < 2R$ and $d(x_2, x'_0) < 2^2 R$. Continuing in the same way, we obtain a chain (x_0, x_1, \dots, x_k) such that $d(x_{i-1}, x_i) < 2^{i-1} R$ and $d(x_i, x'_0) < 2^i R$, $1 \leq i \leq k$. Since $k \leq R$, this completes the proof for $\ell = 0$.

Case $\ell \geq 1$: Let y_0, \dots, y_{k-1} with $y_i \in Y_{\alpha^i(v)}$ and $y'_0, \dots, y'_{\ell-1}$ with $y'_j \in Y_{\alpha^j(v')}$ be sequences such that the expression whose infimum is taken in (13) is less than $R - k - \ell$. Let $z = f_{\alpha^{k-1}(v)}(y_{k-1})$ and $z' = f_{\alpha^{\ell-1}(v')}(y'_{\ell-1})$. Then we have $d(x_0, z) + d(z, z') + d(x'_0, z') < R$, so that both $d(x_0, z)$ and $d(x'_0, z)$ are less than R . The proof is completed by applying the first part of the proof to the pairs x_0, z and x'_0, z and noting that

$$d(x_k, x'_\ell) \leq d(x_k, z) + d(x'_\ell, z) < 2^k R + 2^\ell R \leq 2^{k+\ell} R \leq 2^R R. \quad \square$$

Given an n -chain $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ starting in X_v , define, for $0 \leq k \leq n-1$, $\delta_k = d(x_k, Y_{\alpha^k(v)})$ and $\theta_k = \delta_0 \vee \dots \vee \delta_k$, where we introduce the notation $a \vee b = \max\{a, b\}$. For future notational convenience define $\theta_{-1} = 0$. Note that if $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ is another n -chain starting in X_v , then

$$(15) \quad d(x_{k-1}, x_k) \leq \delta_{k-1} + 2,$$

$$(16) \quad |\delta_k - \delta'_k| = |d(x_k, Y_{\alpha^k(v)}) - d(x'_k, Y_{\alpha^k(v)})| \leq d(x_k, x'_k), \quad 0 \leq k \leq n-1.$$

Lemma 6.2. *Let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ be n -chains starting in X_v . Then*

$$(17) \quad |d(x_k, x'_k) - d(x_0, x'_0)| \leq 2k(\theta_{k-1} \vee \theta'_{k-1}) + 2k, \quad 0 \leq k \leq n-1.$$

Proof. Since the statement is obvious in the case $k = 0$, we assume $k \geq 1$. Let \bar{x}_k and \bar{x}'_k be as in the definition of n -chains. Observe that

$$\begin{aligned} d(\bar{x}_k, \bar{x}'_k) &\leq d(\bar{x}_k, x_k) + d(\bar{x}'_k, x'_k) + d(x_k, x'_k) \leq d(x_k, x'_k) + 2(\theta_k \vee \theta'_k) + 2, \\ d(\bar{x}_k, \bar{x}'_k) &\geq d(x_k, x'_k) - d(\bar{x}_k, x_k) - d(\bar{x}'_k, x'_k) \geq d(x_k, x'_k) - 2(\theta_k \vee \theta'_k) - 2. \end{aligned}$$

Since $f_{\alpha^k(v)}$ is an isometry we have $d(x_{k+1}, x'_{k+1}) = d(\bar{x}_k, \bar{x}'_k)$, and the lemma follows from these inequalities by induction. \square

Given an n -chain $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $N > 0$, we define, for $0 \leq k \leq n-1$,

$$(18) \quad a_k = \left(1 - \frac{\ln(1 + \theta_k)}{N}\right)_+,$$

$$(19) \quad c_k = \sqrt{a_0 \cdots a_{k-1} (1 + (n - k - 1)(1 - a_k))},$$

where $a_+ = \max\{a, 0\}$. Note that $c_0 = \sqrt{1 + (n-1)(1-a_0)}$. One checks immediately that $a_0 \geq a_1 \geq \cdots \geq a_{n-1}$,

$$(20) \quad c_0^2 + \cdots + c_{n-1}^2 = n, \quad 0 \leq c_{n-1} \leq 1,$$

$$(21) \quad \delta_k \geq e^N - 1 \implies \theta_k \geq e^N - 1 \implies a_k = 0 \implies c_{k+1} = \cdots = c_{n-1} = 0.$$

The coefficients c_k were introduced by Tu [19] (actually we work with the square root of Tu's coefficients). Since the maps f_v are isometries, it is possible in our case to define a_k explicitly as in (18). Both a_k and c_k should be regarded as functions $a_k^{\mathbf{x}}$ and $c_k^{\mathbf{x}}$ of n -chains $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$. We will often write a'_k and c'_k instead of $a_k^{\mathbf{x}'}$ and $c_k^{\mathbf{x}'}$.

Lemma 6.3. *Let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ be n -chains starting in X_v . If $\omega = \max_{0 \leq i \leq n-1} |a_i - a'_i|$, then*

$$\begin{aligned} (i) \quad &|c_k^2 - c'^2_k| \leq \left(\frac{n+1}{2}\right)^2 \omega, \\ (ii) \quad &|c_k - c'_k| \leq \frac{n+1}{\sqrt{2}} \sqrt{\omega}, \text{ and} \\ (iii) \quad &\sum_{k=0}^{n-1} \frac{1}{n} |c_k - c'_k|^2 \leq \frac{(n+1)^2}{2} \omega. \end{aligned}$$

Proof. This is an exercise. For (i) \implies (ii) use the inequality $|\sqrt{a} - \sqrt{b}| \leq \sqrt{2|a - b|}$. \square

The following continuity property of the coefficients a_k is a minor variation of a formula in Tu's paper [19, formula 9.2].

Lemma 6.4. *Let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ be n -chains starting in X_v , and assume $d(x_0, x'_0) \leq n$. Then*

$$(22) \quad \max_{0 \leq k \leq n-1} |a_k - a'_k| \leq \frac{7n^2}{N}.$$

Proof. Denote $\lambda = \ln 7n$. We are going to show that

$$(23) \quad \begin{aligned} |a_0 - a'_0| &\leq \frac{n}{N}, \\ |a_k - a'_k| &\leq |a_{k-1} - a'_{k-1}| + \frac{\lambda}{N}. \end{aligned}$$

From these we immediately obtain $|a_k - a'_k| \leq \frac{n(1+\lambda)}{N} \leq \frac{7n^2}{N}$, concluding the proof. Of these inequalities, the first is straightforward. Indeed, using (16) and the property that the map $t \mapsto (1 - \frac{1}{N} \ln(1+t))_+$ is $\frac{1}{N}$ -Lipschitz, we have

$$|a_0 - a'_0| \leq \frac{1}{N} |\delta_0 - \delta'_0| \leq \frac{1}{N} d(x_0, x'_0) \leq \frac{n}{N}.$$

To prove the second inequality, denote $f(t) = 4nt + 6n$. According to (16) and Lemma 6.2, we have

$$(24) \quad |\delta_k - \delta'_k| \leq d(x_k, x'_k) \leq \frac{1}{2} f(\theta_{k-1} \vee \theta'_{k-1}).$$

Denote $\psi(t) = \ln(1+t)$, so that for all $t \geq 0$ we have

$$(25) \quad \psi(t) \leq \psi(f(t)) \leq \psi(t) + \lambda.$$

Case $\theta_k \vee \theta'_k \leq f(\theta_{k-1} \vee \theta'_{k-1})$: By symmetry we may assume that $\theta_{k-1} \geq \theta'_{k-1}$. Thus $\theta'_{k-1} \leq \theta'_k \leq f(\theta_{k-1})$ and $\theta'_{k-1} \leq \theta_{k-1} \leq \theta_k \leq f(\theta_{k-1})$. Using (25), we obtain

$$\begin{aligned} \psi(\theta'_{k-1}) &\leq \psi(\theta'_k) \leq \psi(f(\theta_{k-1})) \leq \psi(\theta_{k-1}) + \lambda, \\ \psi(\theta'_{k-1}) &\leq \psi(\theta_k) \leq \psi(f(\theta_{k-1})) \leq \psi(\theta_{k-1}) + \lambda. \end{aligned}$$

Now, (23) follows immediately from these inequalities, together with the property that for real numbers $a \leq s, t \leq b + \lambda$ we have

$$|(1 - s/N)_+ - (1 - t/N)_+| \leq |(1 - b/N)_+ - (1 - a/N)_+| + \lambda/N.$$

Case $\theta_k \vee \theta'_k \geq f(\theta_{k-1} \vee \theta'_{k-1})$: By symmetry we may assume that $\theta_k \geq \theta'_k$. Then $\theta_k \geq f(\theta_{k-1}) \geq \theta_{k-1}$, and hence $\theta_k = \delta_k$. Therefore, using (24),

$$\theta_k \leq |\theta_k - \theta'_k| + \theta'_k \leq |\delta_k - \delta'_k| + \theta'_k \leq \frac{1}{2} f(\theta_{k-1} \vee \theta'_{k-1}) + \theta'_k \leq \frac{1}{2} \theta_k + \theta'_k.$$

Hence $\theta'_k \leq \theta_k \leq 2\theta'_k$. We obtain

$$\psi(\theta'_k) \leq \psi(\theta_k) \leq \psi(\theta'_k) + \ln 2.$$

From this inequality and the property that the map $t \mapsto (1 - t/N)_+$ is $\frac{1}{N}$ -Lipschitz, we obtain $|a_k - a'_k| \leq \frac{\ln 2}{N} \leq \frac{\lambda}{N}$, which implies (23). \square

Definition. Given $R > 0$ and $\varepsilon > 0$ choose and fix $n \in \mathbb{N}$ such that

$$(26) \quad \left(\frac{\ln n + 9}{4n} \right)^{\frac{1}{2}} < \frac{\varepsilon}{3(R+1)}, \quad n > 2^R R,$$

and $N \in \mathbb{N}$ such that

$$(27) \quad \frac{6n(n+1)}{\sqrt{N}} < \frac{\varepsilon}{3(R+1)}.$$

Having done so, apply the fundamental criterion for uniform embeddability, Proposition 2.1, to choose and fix a family $(\xi_x)_{x \in X}$ of unit vectors in a Hilbert space $\mathcal{H}_X = \bigoplus_{v \in V} \mathcal{H}_v$ with $\xi_x \in \mathcal{H}_v$ if $x \in X_v$ and such that

$$(28) \quad \sup\{\|\xi_y - \xi_{y'}\| : d(y, y') \leq 3ne^N, y, y' \in X_v, v \in V\} < \frac{\varepsilon}{3(R+1)},$$

$$(29) \quad \lim_{S \rightarrow \infty} \sup\{|\langle \xi_y, \xi_{y'} \rangle| : d(y, y') \geq S, y, y' \in X_v, v \in V\} = 0.$$

(See the remarks after Theorem 5.3 for comments on why this is possible.) Finally, for every n -chain $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ in X define the unit vector $\eta^{\mathbf{x}} \in \mathcal{H}_X$ by

$$(30) \qquad \eta^{\mathbf{x}} = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} c_k \xi_{x_k},$$

where, of course, the c_k 's are defined according to (19) and depend on the chain \mathbf{x} .

Lemma 6.5. *Let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ be n -chains starting in X_v . If $d(x_0, x'_0) \leq n$, then $\|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| \leq \frac{2\varepsilon}{3(R+1)}$.*

Proof. Let $I = \{k : c_k \neq 0, c'_k \neq 0\}$ and $J = \{k : c_k = 0, c'_k \neq 0\}$. If $1 \leq k \in I$, then $a_{k-1} = (1 - \frac{1}{N} \ln(1 + \theta_{k-1}))_+ \neq 0$; hence $\theta_{k-1} \leq e^N - 1$. Similarly, $\theta'_{k-1} \leq e^N - 1$. Hence from (17) we have

$$(31) \qquad d(x_k, x'_k) \leq 2n(e^N - 1) + 2n + d(x_0, x'_0) \leq 3ne^N.$$

Consider

$$\begin{aligned} \|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| &= \left\| \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} c_k \xi_{x_k} - \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} c'_k \xi_{x'_k} \right\| \\ &\leq \left\{ \left\| \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (c_k - c'_k) \xi_{x_k} \right\| + \left\| \frac{1}{\sqrt{n}} \sum_{k \in J} (c'_k - c_k) (\xi_{x_k} - \xi_{x'_k}) \right\| \right. \\ &\qquad \left. + \left\| \frac{1}{\sqrt{n}} \sum_{k \in I} c'_k (\xi_{x_k} - \xi_{x'_k}) \right\| \right\}. \end{aligned}$$

Observe that the ξ_{x_k} 's and $\xi_{x'_k}$'s are in orthogonal components of \mathcal{H}_X . We bound the third term on the right using (20), (31) and (28) as follows:

$$\begin{aligned} \frac{1}{\sqrt{n}} \left(\sum_{k \in I} (c'_k)^2 \|\xi_{x_k} - \xi_{x'_k}\|^2 \right)^{1/2} &\leq \frac{1}{\sqrt{n}} \left(\sum_{k=0}^{n-1} (c'_k)^2 \right)^{1/2} \cdot \left(\sup_{k \in I} \|\xi_{x_k} - \xi_{x'_k}\| \right) \\ &\leq \sup_{\Delta} \|\xi_y - \xi_{y'}\| \leq \frac{\varepsilon}{3(R+1)}, \end{aligned}$$

where $\Delta = \{(y, y') : d(y, y') \leq 3ne^N, y, y' \in X_v, v \in V\}$. We bound the sum of the first two terms on the right, by

$$\begin{aligned} 3 \left(\sum_{k=0}^{n-1} \frac{1}{n} |c_k - c'_k|^2 \right)^{\frac{1}{2}} &\leq \frac{3(n+1)}{\sqrt{2}} \max_{0 \leq i \leq n-1} |a_i - a'_i|^{\frac{1}{2}} \\ &\leq \frac{3(n+1)}{\sqrt{2}} \left(\frac{7n^2}{N} \right)^{\frac{1}{2}} \leq \frac{\varepsilon}{3(R+1)}, \end{aligned}$$

where the inequalities are from Lemma 6.3, (22) and (27), respectively. Combining these observations, we obtain the result. \square

Lemma 6.6. *Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{x}' = (x_0, x'_1, \dots, x'_{n-1})$ be n -chains with $x_0 \in X_v$ and $x_1 \in X_{\alpha(v)}$. If (x_0, x_1) is a 2-chain and $d(x_0, x_1) \leq n$, then $\|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| \leq \frac{\varepsilon}{R+1}$.*

Proof. Let $\bar{x}_0 = f_v^{-1}(x_1)$ and set $\bar{\mathbf{x}} = (\bar{x}_0, x_1, \dots, x_{n-1})$. According to the definition (30) of $\eta^{\mathbf{x}}$ we have

$$\eta^{\mathbf{x}} = \eta^{(x_1, x_2, \dots, x_n)} = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{n-1} d_k \xi_{x_k} + d_n \xi_{x_n} \right),$$

$$\eta^{\bar{\mathbf{x}}} = \eta^{(\bar{x}_0, x_1, \dots, x_{n-1})} = \frac{1}{\sqrt{n}} \left(\bar{d}_0 \xi_{\bar{x}_0} + \sum_{k=1}^{n-1} \bar{d}_k \xi_{x_k} \right),$$

where

$$d_k = c_{k-1}(x_1, \dots, x_k) = \sqrt{\bar{a}_1 \cdots \bar{a}_{k-1} (1 + (n-k)(1 - \bar{a}_k))},$$

$$\bar{d}_k = c_k(\bar{x}_0, x_1, \dots, x_k) = \sqrt{\bar{a}_0 \bar{a}_1 \cdots \bar{a}_{k-1} (1 + (n-k-1)(1 - \bar{a}_k))},$$

with $\bar{a}_0 = 1$, $\bar{d}_0 = 1$ and $\bar{a}_k = a_k(\bar{x}_0, x_1, \dots, x_k) = \min\{(1 - \frac{\ln(1+\delta_i)}{N})_+ : 1 \leq i \leq k\}$. For $1 \leq k \leq n-1$ we have either $1 - \bar{a}_k = 0$, hence $d_k - \bar{d}_k = 0$, or

$$d_k - \bar{d}_k = \frac{\sqrt{\bar{a}_1 \cdots \bar{a}_{k-1}} (1 - \bar{a}_k)}{\sqrt{1 + (n-k)(1 - \bar{a}_k)} + \sqrt{1 + (n-k-1)(1 - \bar{a}_k)}}$$

$$\leq \frac{(1 - \bar{a}_k)}{2\sqrt{(n-k)(1 - \bar{a}_k)}} \leq \frac{1}{2\sqrt{n-k}}.$$

Applying the above expressions for $\eta^{\mathbf{x}}$ and $\eta^{\bar{\mathbf{x}}}$ with this inequality, we estimate

$$\|\eta^{\bar{\mathbf{x}}} - \eta^{\mathbf{x}}\| = \frac{1}{\sqrt{n}} \left\| \bar{d}_0 \xi_{\bar{x}_0} + \sum_{k=1}^{n-1} (\bar{d}_k - d_k) \xi_{x_k} - d_n \xi_{x_n} \right\|$$

$$\leq \left(\frac{2}{n} + \frac{1}{n} \sum_{k=1}^{n-1} (d_k - \bar{d}_k)^2 \right)^{\frac{1}{2}}$$

$$\leq \left(\frac{2}{n} + \frac{1}{4n} \left(\frac{1}{n-1} + \frac{1}{n-2} + \cdots + 1 \right) \right)^{\frac{1}{2}}$$

$$\leq \left(\frac{\ln n + 9}{4n} \right)^{\frac{1}{2}} \leq \frac{\varepsilon}{3(R+1)},$$

where the final inequality comes from the choice (26) of n . Apply Lemma 6.5 to the chains \mathbf{x}' and $\bar{\mathbf{x}}$, noting that $d(x_0, \bar{x}_0) = d(x_0, x_1) - 1 \leq n$, and use the previous inequality to conclude that

$$\|\eta^{\mathbf{x}'} - \eta^{\mathbf{x}}\| \leq \|\eta^{\mathbf{x}'} - \eta^{\bar{\mathbf{x}}}\| + \|\eta^{\bar{\mathbf{x}}} - \eta^{\mathbf{x}}\| \leq \frac{2\varepsilon}{3(R+1)} + \frac{\varepsilon}{3(R+1)} = \frac{\varepsilon}{R+1}. \quad \square$$

Lemma 6.7. Let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and $\mathbf{x}' = (x'_0, x'_1, \dots, x'_{n-1})$ be n -chains in X . Then

$$|\langle \eta^{\mathbf{x}}, \eta^{\mathbf{x}'} \rangle| \leq \sup \{ |\langle \xi_y, \xi_{y'} \rangle| : y, y' \in X_v, v \in V, d(y, y') \geq d(x_0, x'_0) - 6ne^N \}.$$

Proof. Let k and ℓ be as in (13). Considering symmetry, we assume that $k \geq \ell$. Also, if $k \geq n$, then $\langle \eta^{\mathbf{x}}, \eta^{\mathbf{x}'} \rangle = 0$; so we assume that $k < n$. With these assumptions

$$(32) \quad \langle \eta^{\mathbf{x}}, \eta^{\mathbf{x}'} \rangle = \frac{1}{n} \sum_{i=0}^{n-k-1} c_{k+i} c'_{\ell+i} \langle \xi_{x_{k+i}}, \xi_{x'_{\ell+i}} \rangle.$$

Making use of (21), we conclude that if $\theta_{k-1} \vee \theta'_{\ell-1} \geq e^N - 1$, then all the products $c_{k+i}c'_{\ell+i}$ in (32) are zero, and we are done. Thus we assume that $\theta_{k-1} \vee \theta'_{\ell-1} \leq e^N - 1$. Let m be the largest number $0 \leq m \leq n - k - 1$ with the property that $\theta_{k+m-1} \vee \theta'_{\ell+m-1} \leq e^N - 1$. By (17),

$$d(x_{k+i}, x'_{\ell+i}) \geq d(x_k, x'_\ell) - 2ne^N$$

for all $0 \leq i \leq m$. On the other hand, using (15),

$$\begin{aligned} d(x_k, x'_\ell) &\geq d(x_0, x'_0) - \sum_{i=0}^{k-1} d(x_i, x_{i+1}) - \sum_{j=0}^{\ell-1} d(x'_j, x'_{j+1}) \\ &\geq d(x_0, x'_0) - \sum_{i=0}^{k-1} (\delta_i + 2) - \sum_{j=0}^{\ell-1} (\delta'_j + 2) \\ &\geq d(x_0, x'_0) - (k + \ell)(2(\theta_{k-1} \vee \theta'_{\ell-1}) + 2) \\ &\geq d(x_0, x'_0) - 4ne^N. \end{aligned}$$

Combining these two inequalities, we obtain

$$d(x_{k+i}, x'_{\ell+i}) \geq d(x_0, x'_0) - 6ne^N, \quad \text{for all } 0 \leq i \leq m.$$

Finally, arguing again on the basis of (21) as above, we conclude that the terms in (32) for $i > m$ are zero. Thus, applying (20) and the previous inequality, we see that

$$|\langle \eta^{\mathbf{x}}, \eta^{\mathbf{x}'} \rangle| \leq \left(\frac{1}{n} \sum_{i=0}^m c_{k+i}c'_{\ell+i} \right) \cdot \sup_{\Omega} |\langle \xi_y, \xi_{y'} \rangle| \leq \sup_{\Omega} |\langle \xi_y, \xi_{y'} \rangle|,$$

where $\Omega = \{(y, y') : y, y' \in X_v, v \in V, d(y, y') \geq d(x_0, x'_0) - 6ne^N\}$. \square

Proposition 6.8. *Given $R > 0$ and $\varepsilon > 0$, let n , N , and $(\xi_x)_{x \in X}$ be constructed as in the definition. For each $x_0 \in X$, choose and fix an n -chain $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ and consider the corresponding vector $\eta^{\mathbf{x}} = \eta^{(x_0, x_1, \dots, x_{n-1})}$. Then*

$$(33) \quad \sup\{\|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| : d(x_0, x'_0) < R\} \leq \varepsilon,$$

$$(34) \quad |\langle \eta^{\mathbf{x}}, \eta^{\mathbf{x}'} \rangle| \leq \sup\{|\langle \xi_y, \xi_{y'} \rangle| : d(y, y') \geq d(x_0, x'_0) - 6ne^N, y, y' \in X_v, v \in V\}.$$

Proof. The support condition (34) was proven in Lemma 6.7. For the convergence condition (33) we show that for any $x_0, x'_0 \in X$ with $d(x_0, x'_0) < R$ and any two n -chains \mathbf{x} and \mathbf{x}' starting at x_0 and x'_0 , respectively, we have $\|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| \leq \varepsilon$. Let k and ℓ be as in (13) and (14). Let $\mathbf{x}(i)$ and $\mathbf{x}'(j)$ be n -chains whose initial elements are the points x_i, x'_j given by Lemma 6.1, $0 \leq i \leq k$, $0 \leq j \leq \ell$. Applying Lemmas 6.5 and 6.6 repeatedly (note that since $n > 2^R R$ this is possible by Lemma 6.1), with the convention that all empty sums are zero, we have

$$\begin{aligned} \|\eta^{\mathbf{x}} - \eta^{\mathbf{x}'}\| &\leq \sum_{i=0}^{k-1} \|\eta^{\mathbf{x}(i)} - \eta^{\mathbf{x}(i+1)}\| + \|\eta^{\mathbf{x}(k)} - \eta^{\mathbf{x}'(\ell)}\| + \sum_{j=0}^{\ell-1} \|\eta^{\mathbf{x}'(j)} - \eta^{\mathbf{x}'(j+1)}\| \\ &\leq \frac{(k + \ell + 1)\varepsilon}{R + 1} \leq \varepsilon. \end{aligned}$$

\square

Conclusion of the proof of Theorem 5.3. Given $R > 0$ and $\varepsilon > 0$, let $x_0 \mapsto \eta^{(x_0, x_1, \dots, x_{n-1})}$ be a map constructed as in Proposition 6.8. It satisfies the convergence and support conditions of Proposition 2.1, as shown by (33) and by (34) in conjunction with (29). \square

Let us note that Proposition 6.8 also reproves the following fact.

Theorem 6.9 ([8], [19]). *Let A and B be countable discrete groups and let C be a common subgroup. If both A and B are exact, then the amalgamated free product $A *_C B$ is exact.*

Proof. Given $R > 0$ and $\varepsilon > 0$, let $x_0 \mapsto \eta^{(x_0, x_1, \dots, x_{n-1})}$ be a map constructed as in Proposition 6.8. If the family $(\xi_x)_{x \in X}$ satisfies the support condition (ii) of Proposition 2.6 uniformly with respect to $v \in V$, then it follows from (34) that $\langle \eta^x, \eta^{x'} \rangle = 0$ whenever $d(x_0, x'_0) \geq S + 6ne^N$. In particular, the map $x_0 \mapsto \eta^{(x_0, x_1, \dots, x_{n-1})}$ satisfies the convergence and support conditions of Proposition 2.6, and $A *_C B$ is exact by Corollary 2.8. \square

7. APPENDIX

We collect several elementary results on the construction of metrics required for our treatment of amalgamated free products. Let X be a set. A *partial metric* on X is an integer-valued function $\widehat{d} : \mathcal{D} \rightarrow \mathbb{R}_+$ defined on a domain $\mathcal{D} \subset X \times X$ and satisfying, for all $x, y \in X$,

- (i) $\widehat{d}(x, y) = \widehat{d}(y, x)$, and
- (ii) $\widehat{d}(x, y) = 0$ if and only if $x = y$.

In these statements it is assumed that all relevant pairs belong to the domain \mathcal{D} of \widehat{d} and that the domain \mathcal{D} is symmetric and contains the diagonal.

We now associate a metric to a partial metric. The construction is analogous to that of path metrics; intuitively the distance between two points is the length of the shortest path between them. A *path* from x to $y \in X$ is a sequence $x = x_0, x_1, \dots, x_n = y$ such that for every $1 \leq j \leq n$ we have $(x_{j-1}, x_j) \in \mathcal{D}$. The *length* of a path is $\sum_{i=1}^n \widehat{d}(x_{i-1}, x_i)$. The domain \mathcal{D} is *ample* if for every $x, y \in \mathcal{D}$ there exists a path from x to y .

Proposition 7.1. *Let \widehat{d} be a partial metric on X , defined on an ample domain \mathcal{D} . Define, for all $x, y \in X$,*

$$d(x, y) = \inf \{ \text{length of paths from } x \text{ to } y \}.$$

Then d is an integer-valued metric on X . \square

The metric d defined in the proposition is the *metric envelope* of \widehat{d} .

Example. If X is the vertex set of a connected graph and \widehat{d} is the constant 1 on the domain of all pairs (x, y) that represent edges, then d is the path metric; $d(x, y)$ is the smallest number of edges on a path from x to y .

Remark. If $(x, y) \in \mathcal{D}$, then $d(x, y) = \widehat{d}(x, y)$ if and only if the length of every path from x to y is greater than or equal to $\widehat{d}(x, y)$. If $\mathcal{D} = X \times X$ and \widehat{d} is a metric, then $d = \widehat{d}$.

Proposition 7.2. *Let \widehat{d} be a partial metric on X and let $\varphi : X \rightarrow X$ be a bijection with the property that*

- (i) $(x, y) \in \mathcal{D}$ if and only if $(\varphi(x), \varphi(y)) \in \mathcal{D}$, and
- (ii) $\widehat{d}(\varphi(x), \varphi(y)) = \widehat{d}(x, y)$ if $(x, y) \in \mathcal{D}$.

Then φ is an isometry for the metric envelope.

Proof. It suffices to show that

$$d(\varphi(x), \varphi(y)) \leq d(x, y), \quad \text{for all } x, y \in X.$$

Let $\varepsilon > 0$ be given. Let (x_0, \dots, x_n) be a path in X from x to y such that $\sum_{j=1}^n \widehat{d}(x_{j-1}, x_j) < d(x, y) + \varepsilon$. Then $(\varphi(x_0), \dots, \varphi(x_n))$ is a path from $\varphi(x)$ to $\varphi(y)$, and we have

$$d(\varphi(x), \varphi(y)) \leq \sum_{i=1}^n \widehat{d}(\varphi(x_{i-1}), \varphi(x_i)) = \sum_{j=1}^n \widehat{d}(x_{j-1}, x_j) < d(x, y) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we are done. \square

REFERENCES

- [1] J. Alonso and M. Bridson, *Semihyperbolic groups*, Proc. London Math. Soc. **70** (1995), 56–114. MR **95j**:20033
- [2] C. Anantharaman-Delaroche, *Amenability and exactness for dynamical systems and their C^* -algebras*, Trans. Amer. Math. Soc. **354** (2002), 4153–4178.
- [3] C. Anantharaman-Delaroche and J. Renault, *Amenable groupoids*, with a foreword by Georges Skandalis and Appendix B by E. Germain, Monographies de L'Enseignement Math., vol. 36, L'Enseignement Math., Geneva, 2000. MR **2001m**:22005
- [4] G. Baumslag, *Topics in combinatorial group theory*, ETH Lectures in Mathematics, Birkhäuser-Verlag, Basel, 1993. MR **94j**:20034
- [5] X. Chen, M. Dadarlat, E. Guentner, and G. Yu, *Uniform embeddings into Hilbert space and free products of groups*, to appear in J. Funct. Anal.
- [6] P. Cherix, M. Cowling, P. Jolissaint, P. Julg, and A. Valette, *Groups with the Haagerup property. Gromov's α -T-menability*, Progress in Mathematics, vol. 197, Birkhäuser-Verlag, Basel, 2001. MR **2002h**:22007
- [7] A. Dranishnikov, G. Gong, V. Lafforgue, and G. Yu, *Uniform embeddings into Hilbert space and a question of Gromov*, Canad. Math. Bull. **45** (2002), 60–70. MR **2003a**:57043
- [8] K. Dykema, *Exactness of reduced amalgamated free product C^* -algebras*, Preprint, 1999.
- [9] ———, *Free products of exact groups, C^* -algebras* (Münster, 1999) (J. Cuntz and S. Echterhoff, eds.), Springer-Verlag, Berlin, 2000, pp. 61–70. MR **2001m**:46133
- [10] S. Ferry, A. Ranicki, and J. Rosenberg (eds.), *Novikov conjectures, index theorems and rigidity*, London Mathematical Society Lecture Notes, nos. 226, 227, Cambridge University Press, 1995. MR **96m**:57002; MR **96m**:57003
- [11] S. Gersten, *Bounded cocycles and combings of groups*, unpublished manuscript, version 5.5, 1991, cf. Internat. J. Algebra Comput. **2** (1992), 307–326. MR **93i**:20029
- [12] M. Gromov, *Asymptotic invariants of infinite groups*, London Mathematical Society Lecture Notes, no. 182, pp. 1–295, Cambridge University Press, Cambridge, 1993. MR **95m**:20041
- [13] E. Guentner and J. Kaminker, *Exactness and the Novikov conjecture*, Topology **41** (2002), no. 2, 411–418. MR **2003e**:46097a
- [14] E. Kirchberg and S. Wassermann, *Exact groups and continuous bundles of C^* -algebras*, Mathematische Annalen **315** (1999), 169–203. MR **2000i**:46050
- [15] ———, *Permanence properties of C^* -exact groups*, Documenta Mathematica **4** (1999), 513–558 (electronic). MR **2001i**:46089
- [16] N. Ozawa, *Amenable actions and exactness for discrete groups*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 8, 691–695. MR **2001g**:22007
- [17] J. P. Serre, *Trees*, Springer-Verlag, New York, 1980, Translation from French of “Arbres, Amalgames, SL_2 ”, Astérisque no. 46. MR **82c**:20083
- [18] G. Skandalis, J. L. Tu, and G. Yu, *The coarse Baum-Connes conjecture and groupoids*, Topology **41** (2002), 807–834. MR **2003c**:58020
- [19] J. L. Tu, *Remarks on Yu's Property A for discrete metric spaces and groups*, Bull. Soc. Math. France **129** (2001), 115–139. MR **2002j**:58038

- [20] G. Yu, *The Coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert space*, *Inventiones Math.* **139** (2000), 201–240. MR 2000j:19005

DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, 1395 MATHEMATICAL SCIENCES BUILDING, WEST LAFAYETTE, INDIANA 47907-1395

E-mail address: mdd@math.purdue.edu

MATHEMATICS DEPARTMENT, UNIVERSITY OF HAWAII, MANOA, 2565 MCCARTHY MALL, HONOLULU, HAWAII 96822

E-mail address: erik@math.hawaii.edu

VITALI COVERING THEOREM IN HILBERT SPACE

JAROSLAV TIŠER

ABSTRACT. It is shown that the statement of the Vitali Covering Theorem does not hold for a certain class of measures in a Hilbert space. This class contains all infinite-dimensional Gaussian measures.

1. INTRODUCTION

We start with recalling the statement of the classical covering theorem due to G. Vitali, [9].

Theorem 1. *Let $A \subset \mathbb{R}^n$ be a set. Assume that for every $x \in A$ there is a sequence $(B[x, r_k(x)])_k$ of closed balls centred at x and with radii $r_k(x)$ such that $\lim_{k \rightarrow \infty} r_k(x) = 0$. Then there is an at most countable family of disjoint balls, $\{B[x_i, r_{k_i}(x_i)] \mid i \in \mathbb{N}\}$, such that*

$$\mathcal{L}_n \left(A \setminus \bigcup_{i \in \mathbb{N}} B[x_i, r_{k_i}(x_i)] \right) = 0.$$

The balls in the original paper were considered with respect to the norm $\|\cdot\|_\infty$. In fact, the statement of the theorem above holds true for balls in any equivalent norm in \mathbb{R}^n .

Since the time of Vitali there appeared many generalizations of the statement in various directions. To mention at least one of them, now already classical, we have to point out the version based on the Besicovitch Covering Theorem. It extends the statement from Lebesgue measure to any σ -finite measure on \mathbb{R}^n ; see e.g. de Guzmán [1].

Our aim is to study what happens if we replace the n -dimensional Euclidean space \mathbb{R}^n by an infinite-dimensional Hilbert space. The first result of this type is due to D. Preiss, [4]. He gave an example of a Gaussian measure on a separable Hilbert space for which the covering theorem fails to hold.

One of the most important consequences of the Vitali Covering Theorem is the so-called Differentiation theorem. The original version goes back to H. Lebesgue. Employing the above-mentioned generalization of the covering theorem, one has the following form of the Differentiation theorem. Here, and also in the sequel, $B[x, r]$ denotes the closed ball with center x and radius $r > 0$.

Received by the editors May 15, 2002 and, in revised form, January 24, 2003.

2000 *Mathematics Subject Classification.* Primary 28A50, 46G99.

Key words and phrases. Vitali Covering Theorem, Gaussian measure, packing density.

The author was supported by grants GA ČR 201/98/1153 and J04/98/210000010.

Differentiation Theorem 1. *Let μ be a locally finite measure on \mathbb{R}^n and let $f \in L^1_{\text{loc}}(\mu)$. Then*

$$(1) \quad \lim_{r \rightarrow 0} \frac{1}{\mu B[x, r]} \int_{B[x, r]} f \, d\mu = f(x) \quad \mu - a.e.$$

The negative result of D. Preiss [4] was later strengthened in [5] by constructing a bounded function and a Gaussian measure on a Hilbert space such that (1) does not hold. Moreover, in [6] the same author obtained a Gaussian measure γ together with the integrable function $f \in L^1(\gamma)$ such that the means of f over the balls in (1) tend to infinity uniformly with respect to x .

On the other hand, J. Tišer [8] has shown the validity of (1) for some class of Gaussian measures on a Hilbert space and for all L^p functions, $1 < p < \infty$. This result could indicate that there is a chance for having the Vitali Covering theorem at least for some infinite-dimensional Gaussian measures. However, Theorem 1 below makes clear that it is not the case, and that Preiss' example [4] was not accidental from this point of view.

Before stating Theorem 1 we recall the concept of a Vitali system.

Definition. Let $A \subset X$ be a subset of a metric space X . A family

$$\mathcal{V} \subset \{B[x, r] \mid x \in A, r > 0\}$$

is called the Vitali system on A if for every $x \in A$ and for every $\varepsilon > 0$ the system \mathcal{V} contains a ball $B[x, r]$ with $r \leq \varepsilon$.

Theorem 1. *Let H be a separable Hilbert space and let γ be a Gaussian measure with $\dim \text{spt} \gamma = \infty$. Then for every $\varepsilon > 0$ there exists a Vitali system \mathcal{V} on $\text{spt} \gamma$ such that any disjoint subfamily $\mathcal{S} \subset \mathcal{V}$ satisfies*

$$\gamma\left(\bigcup \mathcal{S}\right) \leq \varepsilon, \quad \text{i.e.,} \quad \gamma\left(\text{spt} \gamma \setminus \bigcup \mathcal{S}\right) \geq 1 - \varepsilon.$$

Theorem 1 is an easy consequence of the following Proposition 1, which is formulated for more general measures than the Gaussian ones. We make some comments on the other consequences of Proposition 1 at the end of this section. First, however, we shall introduce some notions and notation.

The symbol $\text{spt} \mu$ will denote the support of a measure μ . The projection μ_U of the measure μ onto a closed subspace U of the Hilbert space H is defined by the formula

$$\mu_U A = \mu \pi_U^{-1}(A),$$

where $\pi_U: H \rightarrow U$ denotes the projection and $A \subset U$ is any Borel set in U . If $U \subset H$ is a finite-dimensional subspace, then we shall denote by \mathcal{L}_U the corresponding $\dim U$ -dimensional Lebesgue measure.

We shall also mention some basic facts concerning Gaussian measures.

Definition. A probability measure ν on the real line \mathbb{R} is called a Gaussian measure, if either ν is the Dirac measure supported at 0, or it has the Radon-Nikodým derivative with respect to the Lebesgue measure of the form

$$\frac{d\nu}{d\mathcal{L}_1} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

for some $\sigma > 0$. A Borel probability γ on a separable Hilbert space H is called a Gaussian measure if every projection of γ onto a one-dimensional subspace is a Gaussian measure.

We consider the Dirac measure to be a Gaussian measure only for convenience. It enables us to include among the Gaussian measures also the measures that are supported by a proper subspace of the Hilbert space H .

Let γ be a Gaussian measure on H . The covariance operator $S_\gamma: H \rightarrow H$ is defined by

$$\langle S_\gamma x, y \rangle = \int_H \langle x, h \rangle \langle y, h \rangle d\gamma(h), \quad x, y \in H.$$

The operator S_γ is always nonnegative ($\langle S_\gamma x, x \rangle \geq 0$), selfadjoint and nuclear; see e.g. [3]. If $\text{spt} \gamma = H$, the covariance operator is even positive definite. In that case the eigenvectors of S_γ form an orthonormal basis (e_n) of H with the following nice property: If γ_n is the projection of γ onto the line spanned by e_n , then

$$(2) \quad \gamma = \prod_n \gamma_n.$$

Such representation of γ as a countable product will be useful.

Definition. Let $r > 0$. The symbol $\mathfrak{B}(r)$ denotes the set of all disjoint families of closed balls in H of radius $r > 0$,

$$\mathfrak{B}(r) = \{ \mathfrak{B} \mid \mathfrak{B} \text{ is a disjoint family of balls of radius } r \}.$$

Proposition 1. Let H be a separable Hilbert space and let μ be a finite Borel measure on H with the following property: For every $n \in \mathbb{N}$ there is a finite-dimensional subspace $U \subset H$ such that

- (i) $\dim U \geq n$,
- (ii) μ_U is absolutely continuous with respect to the Lebesgue measure \mathcal{L}_U on U ,
- (iii) $\mu \leq \mu_U \times \mu_{U^\perp}$.

Then

$$\lim_{r \rightarrow 0} \sup \left\{ \mu \left(\bigcup \mathfrak{B} \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} = 0.$$

Proof of Theorem 1. Without loss of generality, we may obviously assume that $\text{spt} \gamma = H$. If we recall the representation (2) of a Gaussian measure as a countable product of one-dimensional Gaussian measures, then we see that the conditions (i) – (iii) of Proposition 1 are satisfied. Indeed, let (e_n) be the orthonormal basis of H consisting of the eigenvectors of the covariance operator S_γ . Then for any $n \in \mathbb{N}$ we put $U = \text{span}\{e_1, \dots, e_n\}$. The conditions (i) and (ii) are obviously true and in the condition (iii) we even obtain equality.

Let $\varepsilon > 0$ be given. By Proposition 1, there is a decreasing sequence of numbers $r_k \searrow 0$ such that

$$(3) \quad \sup \left\{ \gamma \left(\bigcup \mathfrak{B} \right) \mid \mathfrak{B} \in \mathfrak{B}(r_k) \right\} \leq \frac{\varepsilon}{2^k}.$$

We define the following Vitali system:

$$\mathcal{V} = \{ B[x, r_k] \mid x \in H, k \in \mathbb{N} \}.$$

Let $\mathcal{S} \subset \mathcal{V}$ be any disjoint subfamily. Then

$$\mathcal{S} = \bigcup_{k \in \mathbb{N}} \mathcal{S}_k, \quad \mathcal{S}_k = \{ B \in \mathcal{S} \mid \text{radius}(B) = r_k \}.$$

Now, by using (3),

$$\gamma\left(\bigcup \mathcal{S}\right) = \sum_{k=1}^{\infty} \gamma\left(\bigcup \mathcal{S}_k\right) \leq \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon. \quad \square$$

Remark. Note that the finite-dimensional subspaces $U \subset H$ from Proposition 1 need not be nested. Also, if we choose for any $n \in \mathbb{N}$ the corresponding subspace U_n with the properties (i)–(iii), then the linear span of $\{U_n \mid n \in \mathbb{N}\}$ need not be dense in H . Hence the conclusion of Proposition 1, and consequently non-validity of the Vitali Covering theorem, can be obtained e.g. for the following type of measures: Let

$$H = H_0 \oplus H_0^\perp$$

be an orthogonal decomposition of H such that $\dim H_0 = \infty$. Let (μ_n) be any sequence of absolutely continuous probability measures on \mathbb{R} . We consider the measure

$$\mu = \prod_n \mu_n$$

on the space $\mathbb{R}^\mathbb{N}$. Since $H_0 \approx \ell^2 \subset \mathbb{R}^\mathbb{N}$, by the 0–1 law there are only two possibilities: either $\mu H_0 = 0$ or $\mu H_0 = 1$. Assume the latter. In that case for arbitrary finite measure ν on H_0^\perp the product $\mu \times \nu$ on H is an example of a measure satisfying the assumptions of Proposition 1.

2. LEMMATA

Let $U \subset H$ be a closed subspace of the Hilbert space H , and let \mathfrak{B} be a family of disjoint closed balls in H of radius r , $\mathfrak{B} \in \mathfrak{B}(r)$. We denote by \mathfrak{B}_U the family

$$\mathfrak{B}_U = \{U \cap B \mid B \in \mathfrak{B}\}.$$

Obviously, \mathfrak{B}_U is a disjoint family of closed balls in U of radii at most r .

The first Lemma establishes one simple geometrical relationship among the balls in \mathfrak{B}_U .

Lemma 1. *Let $U \subset H$ be a subspace of a separable Hilbert space H , and let $\mathfrak{B} \in \mathfrak{B}(1)$. Let $B[u_1, r_1]$ and $B[u_2, r_2]$ be two different balls from \mathfrak{B}_U . If either $2r_1 \leq r_2$ or $2r_2 \leq r_1$, then*

$$\|u_1 - u_2\| \geq r_1 + r_2 + \frac{1}{2}(\sqrt{10} - 3) \max\{r_1, r_2\}.$$

Proof. Since the balls $B[u_1, r_1]$ and $B[u_2, r_2]$ belong to \mathfrak{B}_U , there are two unit balls $B[x_1, 1]$ and $B[x_2, 1] \in \mathfrak{B}$ such that

$$B[u_1, r_1] = U \cap B[x_1, 1] \quad \text{and} \quad B[u_2, r_2] = U \cap B[x_2, 1].$$

Also, since $H = U \oplus U^\perp$, one has

$$x_1 = u_1 + v_1 \quad \text{and} \quad x_2 = u_2 + v_2,$$

where $u_1, u_2 \in U$ and $v_1, v_2 \in U^\perp$. By the disjointness of the balls in \mathfrak{B} it is readily seen that

$$(4) \quad \|u_1 - u_2\|^2 + \|v_1 - v_2\|^2 = \|x_1 - x_2\|^2 \geq 4.$$

Note that $r_1^2 = 1 - \|v_1\|^2$ and $r_2^2 = 1 - \|v_2\|^2$. Using this in the estimate (4) we obtain

$$\begin{aligned}
 \|u_1 - u_2\|^2 &\geq 2 + r_1^2 + r_2^2 + \|v_1\|^2 + \|v_2\|^2 - \|v_1 - v_2\|^2 \\
 (5) \qquad &= 2 + r_1^2 + r_2^2 + 2\langle v_1, v_2 \rangle \geq 2 + r_1^2 + r_2^2 - 2\|v_1\| \|v_2\| \\
 &= 2 + r_1^2 + r_2^2 - 2\sqrt{(1 - r_1^2)(1 - r_2^2)}.
 \end{aligned}$$

Without loss of generality, we may assume that $r_2 \leq r_1$. Then the assumption in Lemma 1 implies that even $2r_2 \leq r_1$. Let $\delta = \frac{1}{2}(\sqrt{10} - 3)$. In order to prove that

$$\|u_1 - u_2\| \geq r_1(1 + \delta) + r_2,$$

we are going to show that

$$\|u_1 - u_2\|^2 - (r_1(1 + \delta) + r_2)^2 \geq 0.$$

To this end, we use the estimate (5):

$$\begin{aligned}
 &\|u_1 - u_2\|^2 - (r_1(1 + \delta) + r_2)^2 \\
 &\geq 2 + r_1^2 + r_2^2 - 2\sqrt{(1 - r_1^2)(1 - r_2^2)} - (r_1(1 + \delta) + r_2)^2 \\
 &= 2 - 2\sqrt{(1 - r_1^2)(1 - r_2^2)} - r_1^2((1 + \delta)^2 - 1) - 2r_1r_2(1 + \delta) \\
 &= g(r_1, r_2).
 \end{aligned}$$

We shall have to find the minimal value of the function $g(r_1, r_2)$ on the set $\{(r_1, r_2) \mid 0 \leq 2r_2 \leq r_1 \leq 1\}$. Some elementary calculation reveals that the function $r_2 \mapsto g(r_1, r_2)$ is nonincreasing on $[0, \frac{1}{2}r_1]$. One more calculation gives that the function $r_1 \mapsto g(r_1, \frac{1}{2}r_1)$ is nondecreasing on $[0, 1]$ provided that

$$(1 + \delta)^2 + (1 + \delta) - 1 \leq \frac{5}{4}.$$

This condition is guaranteed by our choice of δ . Hence the minimal value of $g(r_1, r_2)$ is attained at the point $(0, 0)$ and is equal to 0. This completes the proof. \square

The next lemma estimates the Lebesgue measure of the intersection of two balls in a special position. The symbol $\alpha(n)$ denotes the volume of the unit Euclidean ball in \mathbb{R}^n ,

$$\alpha(n) = \mathcal{L}_n B[0, 1] = \frac{\pi^{n/2}}{\Gamma(1 + n/2)}.$$

Lemma 2. *There is a $\Delta_0 > 0$ such that for any $x \in \mathbb{R}^n$ with $\|x\| = 3$ and $0 < \delta \leq \Delta_0$ we have the following estimate:*

$$\mathcal{L}_n(B[0, 1 + \delta] \cap B[x, 2(1 + 3\delta)]) \leq \alpha(n - 1) 10^{\frac{n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}}.$$

Proof. Let $x = (3, 0, \dots, 0) \in \mathbb{R}^n$. If we write a point $z \in \mathbb{R}^n$ in the form $z = (z_1, z_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$, then the following equations determine the intersection of the spheres $\{y \mid \|y\| = 1 + \delta\} \cap \{y \mid \|y - x\| = 2(1 + 3\delta)\}$:

$$\begin{aligned}
 z_1^2 + \|z_2\|^2 &= (1 + \delta)^2, \\
 (z_1 - 3)^2 + \|z_2\|^2 &= 4(1 + 3\delta)^2.
 \end{aligned}$$

Eliminating $\|z_2\|$, we get $z_1 = \frac{1}{6}(9 + (1+\delta)^2 - 4(1+3\delta)^2)$. Then a simple geometrical observation reveals that

$$\begin{aligned} & \mathcal{L}_n(B[0, 1+\delta] \cap B[x, 2(1+3\delta)]) \\ & \leq 2 \alpha(n-1) \int_{z_1}^{1+\delta} \left((1+\delta)^2 - t^2 \right)^{\frac{n-1}{2}} dt \\ & = 2 \alpha(n-1) (1+\delta)^n \int_{\theta}^1 (1-u^2)^{\frac{n-1}{2}} du, \end{aligned}$$

where $\theta = \frac{z_1}{1+\delta}$. It is clear that $\theta > 0$ for δ small enough. The explicit condition for δ is $\delta \leq \frac{\sqrt{330}-11}{35}$. We estimate the function $(1-u^2)^{\frac{n-1}{2}}$ by its maximal value on the interval $[\theta, 1]$, and we obtain

$$\begin{aligned} & \leq 2 \alpha(n-1) (1+\delta)^n (1-\theta^2)^{\frac{n-1}{2}} (1-\theta) \\ & \leq 2^{\frac{n+1}{2}} \alpha(n-1) (1+\delta)^n (1-\theta)^{\frac{n+1}{2}}. \end{aligned}$$

Since a short calculation gives that $1-\theta \leq 5\delta$ again for small δ ($\delta \leq 2/35$), we get the desired estimate. Finally, we finish the proof by putting $\Delta_0 = \min\{\frac{\sqrt{330}-11}{35}, \frac{2}{35}\} = \frac{2}{35}$. \square

We introduce the following notation. Let $B = B[x, r]$ be a ball. The symbol

$$(1+\delta)B = B[x, (1+\delta)r]$$

denotes the enlarged ball with the *same* center and $(1+\delta)$ times bigger radius. We shall be using both symbols $(1+\delta)B$ and $B[x, (1+\delta)r]$.

The next lemma contains the key estimate needed in the proof of Proposition 1.

Lemma 3. *There is a number $\delta_0 > 0$ such that for every $r > 0$, every family $\mathfrak{B} \in \mathfrak{B}(r)$ of disjoint balls of radius r , and every finite-dimensional subspace $U \subset H$, the following estimate holds:*

$$\mathcal{L}_U\left((1+\delta)B_0 \setminus \bigcup\{(1+\delta)B \mid B \in \mathfrak{B}_U, B \neq B_0\}\right) \geq \frac{1}{2}(1+\delta)^{\dim U} \mathcal{L}_U B_0$$

provided that $0 < \delta \leq \delta_0$ and $B_0 \in \mathfrak{B}_U$.

Proof. Let $B_0 \in \mathfrak{B}_U$ be fixed. Without loss of generality, we assume that B_0 has center at the origin, $B_0 = B[0, r_0]$, say. Let $\delta > 0$ be such that $2\delta < \frac{1}{2}(\sqrt{10}-3)$. Then, by Lemma 1, we see that the ball $(1+\delta)B_0$ is disjoint with

$$\bigcup\{B[x, (1+\delta)r_x] \mid B[x, r_x] \in \mathfrak{B}_U, 2r_x \leq r_0 \text{ or } 2r_0 \leq r_x\}.$$

Accordingly, the only relevant balls in \mathfrak{B}_U that may interfere with the $(1+\delta)B_0$ are those of radii comparable to r_0 . We denote the centres of such balls by

$$C = \left\{x \in U \setminus \{0\} \mid B[x, r_x] \in \mathfrak{B}_U, (1+\delta)B_0 \cap B[x, (1+\delta)r_x] \neq \emptyset\right\}.$$

Note that for the ball $B[x, r_x] \in \mathfrak{B}_U$ with $x \in C$ we have $\frac{1}{2}r_0 \leq r_x \leq 2r_0$.

We have to estimate the measure of $(1+\delta)B_0 \cap \bigcup_{x \in C} B[x, (1+\delta)r_x]$. Since

$$\mathcal{L}_U\left((1+\delta)B_0 \cap \bigcup_{x \in C} B[x, (1+\delta)r_x]\right) \leq \sum_{x \in C} \mathcal{L}_U\left((1+\delta)B_0 \cap B[x, (1+\delta)r_x]\right),$$

we shall look closer at each intersection $(1+\delta)B_0 \cap B[x, (1+\delta)r_x]$.

Let $x \in C$. First note that

$$r_0 + r_x \leq \|x\| \leq (1 + \delta)(r_0 + r_x).$$

Also, $r_x \leq 2r_0$. We show that

$$(6) \quad B[x, (1 + \delta)r_x] \subset B\left[3r_0 \frac{x}{\|x\|}, 2(1 + 3\delta)r_0\right].$$

To see this, let $y \in B[x, (1 + \delta)r_x]$, i.e., $\|y - x\| \leq (1 + \delta)r_x$. Then

$$\begin{aligned} \left\|y - 3r_0 \frac{x}{\|x\|}\right\| &= \left\|y - x + x\left(1 - \frac{3r_0}{\|x\|}\right)\right\| \\ &\leq \|y - x\| + \|x\| \left|1 - \frac{3r_0}{\|x\|}\right| \\ &\leq (1 + \delta)r_x + \left|\|x\| - 3r_0\right|. \end{aligned}$$

If $3r_0 \geq \|x\|$, then the above calculation finishes with

$$\begin{aligned} &\leq (1 + \delta)r_x + 3r_0 - (r_0 + r_x) = 2r_0 + \delta r_x \\ &\leq 2(1 + \delta)r_0 < 2(1 + 3\delta)r_0. \end{aligned}$$

If, on the other hand, $3r_0 \leq \|x\|$, then we proceed as

$$\begin{aligned} &\leq (1 + \delta)r_x + (1 + \delta)(r_0 + r_x) - 3r_0 \\ &\leq 5(1 + \delta)r_0 - 3r_0 < 2(1 + 3\delta)r_0. \end{aligned}$$

It follows immediately from (6) that

$$(7) \quad (1 + \delta)B_0 \cap B[x, (1 + \delta)r_x] \subset (1 + \delta)B_0 \cap B\left[3r_0 \frac{x}{\|x\|}, 2(1 + 3\delta)r_0\right].$$

Now let $n = \dim U$ for short. If, moreover, $\delta \leq \Delta_0$ from Lemma 2, then we obtain the estimate of the intersection on the right-hand side of (7):

$$\begin{aligned} &\mathcal{L}_n\left((1 + \delta)B_0 \cap B[x, (1 + \delta)r_x]\right) \\ &\leq \mathcal{L}_n\left(B[0, (1 + \delta)r_0] \cap B\left[3r_0 \frac{x}{\|x\|}, 2(1 + 3\delta)r_0\right]\right) \\ &= r_0^n \mathcal{L}_n\left(B[0, (1 + \delta)] \cap B\left[3 \frac{x}{\|x\|}, 2(1 + 3\delta)\right]\right) \\ &\leq r_0^n \alpha(n - 1) 10^{\frac{n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}} \\ &= \frac{\alpha(n - 1)}{\alpha(n)} 10^{\frac{n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}} \mathcal{L}_n B[0, r_0]. \end{aligned}$$

Hence

$$\begin{aligned} (8) \quad \mathcal{L}_n\left((1 + \delta)B_0 \cap \bigcup_{x \in C} B[x, (1 + \delta)r_x]\right) \\ \leq \frac{\alpha(n - 1)}{\alpha(n)} 10^{\frac{n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}} \mathcal{L}_n B[0, r_0] \cdot \text{card } C. \end{aligned}$$

What is missing now is some control over the cardinality of the set C . Fortunately, for our purpose we shall not need any hard estimate. The sufficient upper bound

for card C follows from the comparison of certain volumes. To this end, recall that for all $x \in C$,

$$\|x\| \leq (1 + \delta)(r_x + r_0) \quad \text{and} \quad \frac{1}{2}r_0 \leq r_x \leq 2r_0.$$

Hence

$$(9) \quad \bigcup_{x \in C} B[x, r_x] \subset B[0, (5 + 3\delta)r_0].$$

Also, $B[x, r_x] \supset B[x, \frac{1}{2}r_0]$ for $x \in C$. Combining it with (9) we get

$$\mathcal{L}_n B\left[x, \frac{1}{2}r_0\right] \text{ card } C \leq \mathcal{L}_n B[0, (5 + 3\delta)r_0].$$

Thus

$$\text{card } C \leq 10^n \left(1 + \frac{3}{5}\delta\right)^n \leq 10^n (1 + \delta)^n.$$

Using this estimate in (8) we have

$$\begin{aligned} & \mathcal{L}_n \left((1 + \delta)B_0 \cap \bigcup_{x \in C} B[x, (1 + \delta)r_x] \right) \\ & \leq \frac{\alpha(n-1)}{\alpha(n)} 10^{\frac{n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}} \mathcal{L}_n B[0, r_0] 10^n (1 + \delta)^n \\ & = \frac{\alpha(n-1)}{\alpha(n)} 10^{\frac{3n+1}{2}} (1 + \delta)^{2n} \delta^{\frac{n+1}{2}} \mathcal{L}_n B[0, r_0]. \end{aligned}$$

Since $\frac{\alpha(n-1)}{\alpha(n)} \approx \sqrt{n}$ for $n \rightarrow \infty$, there is $\delta_1 > 0$ such that

$$\frac{\alpha(n-1)}{\alpha(n)} 10^{\frac{3n+1}{2}} (1 + \delta)^n \delta^{\frac{n+1}{2}} \leq \frac{1}{2}$$

for all $n \in \mathbb{N}$ and $0 < \delta \leq \delta_1$. With this choice of δ one has

$$(10) \quad \mathcal{L}_n \left((1 + \delta)B_0 \cap \bigcup_{x \in C} B[x, (1 + \delta)r_x] \right) \leq \frac{1}{2} (1 + \delta)^n \mathcal{L}_n B_0.$$

To complete the proof, we put $\delta_0 = \min\{\Delta_0, \delta_1, \frac{1}{4}(\sqrt{10} - 3)\}$. If now $0 < \delta \leq \delta_0$, then by (10),

$$\begin{aligned} & \mathcal{L}_n \left((1 + \delta)B_0 \setminus \bigcup \{ (1 + \delta)B \mid B \in \mathfrak{B}_U, B \neq B_0 \} \right) \\ & = \mathcal{L}_n (1 + \delta)B_0 - \mathcal{L}_n \left((1 + \delta)B_0 \cap \bigcup_{x \in C} B[x, (1 + \delta)r_x] \right) \\ & \geq (1 + \delta)^n \mathcal{L}_n B_0 - \frac{1}{2} (1 + \delta)^n \mathcal{L}_n B_0 = \frac{1}{2} (1 + \delta)^n \mathcal{L}_n B_0 \end{aligned}$$

and the proof is finished. □

We associate with every $B_0 \in \mathfrak{B}_U$ the set

$$D_{B_0} = (1 + \delta)B_0 \setminus \left(B_0 \cup \bigcup \{ (1 + \delta)B \mid B \in \mathfrak{B}_U, B \neq B_0 \} \right).$$

Then, obviously, $\{D_B \mid B \in \mathfrak{B}_U\}$ is the disjoint system of subsets in U . One consequence of Lemma 3 is the following estimate of the measure of D_B .

Corollary 1. Let $\delta_0 > 0$ be as in Lemma 3 and let $U \subset H$ be a finite-dimensional subspace. Then

$$\mathcal{L}_U D_{B_0} \geq \left(\frac{1}{2}(1+\delta)^{\dim U} - 1 \right) \mathcal{L}_U B_0$$

for every $0 < \delta \leq \delta_0$ and every $B_0 \in \mathfrak{B}_U$.

Proof. Since

$$D_{B_0} \cup B_0 = (1+\delta)B_0 \setminus \bigcup \{(1+\delta)B \mid B \in \mathfrak{B}, B \neq B_0\},$$

we obtain by using Lemma 3,

$$\mathcal{L}_U(D_{B_0} \cup B_0) \geq \frac{1}{2}(1+\delta)^{\dim U} \mathcal{L}_U B_0.$$

The sets D_{B_0} and B_0 are disjoint. So $\mathcal{L}_U(D_{B_0} \cup B_0) = \mathcal{L}_U D_{B_0} + \mathcal{L}_U B_0$, and the statement follows by rearrangement. \square

Now we shall estimate the so-called packing density of the family \mathfrak{B}_U in U . Since U is a finite-dimensional subspace of H , we identify it with \mathbb{R}^n , $n = \dim U$. We put

$$Q_k = [-k, k]^n,$$

the n -dimensional cube in U of side $2k$. With this notation we can state the following

Lemma 4. There is $\delta_0 > 0$ such that for every finite-dimensional subspace $U \cong \mathbb{R}^n$ and every $r > 0$,

$$\limsup_{k \rightarrow \infty} \sup \left\{ \frac{\mathcal{L}_n(Q_k \cap \bigcup \mathfrak{B}_U)}{\mathcal{L}_n Q_k} \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1}$$

for any $0 < \delta \leq \delta_0$ and $n \in \mathbb{N}$ with $\frac{1}{2}(1+\delta)^n - 1 > 0$.

Proof. Let $\mathfrak{B} \in \mathfrak{B}(r)$ be arbitrary and let $\delta_0 > 0$ be as in Lemma 3. We denote by \mathcal{R} the family of all balls in \mathfrak{B}_U such that the $(1+\delta)$ enlargement of B is still contained in the cube Q_k ,

$$\mathcal{R} = \{B \in \mathfrak{B}_U \mid (1+\delta)B \subset Q_k\}.$$

Then

$$\begin{aligned} \mathcal{L}_n(Q_k \cap \bigcup \mathfrak{B}_U) &= \sum_{B \in \mathfrak{B}_U} \mathcal{L}_n(Q_k \cap B) \\ (11) \quad &\leq \sum_{B \in \mathcal{R}} \mathcal{L}_n B + \mathcal{L}_n(Q_k \setminus Q_{k-2r(1+\delta)}) \\ &= \sum_{B \in \mathcal{R}} \mathcal{L}_n B + \mathcal{L}_n Q_k \left[1 - \left(1 - \frac{2r(1+\delta)}{k} \right)^n \right] \end{aligned}$$

provided that $k > 2r(1+\delta)$. By Corollary 1,

$$(12) \quad \mathcal{L}_n B \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1} \mathcal{L}_n D_B$$

for n with $\frac{1}{2}(1+\delta)^n - 1 > 0$. Also, $D_B \subset Q_k$ for any $B \in \mathcal{R}$. Since the sets D_B are disjoint for different B 's, we may sum up the estimates in (12) to get

$$(13) \quad \sum_{B \in \mathcal{R}} \mathcal{L}_n B \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1} \sum_{B \in \mathcal{R}} \mathcal{L}_n D_B \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1} \mathcal{L}_n Q_k.$$

Looking back to (11) one has

$$\mathcal{L}_n(Q_k \cap \bigcup \mathfrak{B}_U) \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1} \mathcal{L}_n Q_k + \mathcal{L}_n Q_k \left[1 - \left(1 - \frac{2r(1+\delta)}{k} \right)^n \right].$$

Since the expression on the right-hand side does not depend on \mathfrak{B} , the same estimate holds true also for the supremum over all $\mathfrak{B} \in \mathfrak{B}(r)$. Hence

$$\begin{aligned} \limsup_{k \rightarrow \infty} \sup \left\{ \frac{\mathcal{L}_n(Q_k \cap \bigcup \mathfrak{B}_U)}{\mathcal{L}_n Q_k} \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \\ \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1} + \limsup_{k \rightarrow \infty} \left[1 - \left(1 - \frac{2r(1+\delta)}{k} \right)^n \right] \\ = \frac{1}{\frac{1}{2}(1+\delta)^n - 1} \end{aligned}$$

and the lemma is proved. □

The straightforward reformulation of the statement of Lemma 4 is the following:
For any cube $Q \subset U \cong \mathbb{R}^n$,

$$(14) \qquad \limsup_{r \rightarrow 0} \sup \left\{ \frac{\mathcal{L}_n(Q \cap \bigcup \mathfrak{B}_U)}{\mathcal{L}_n Q} \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \leq \frac{1}{\frac{1}{2}(1+\delta)^n - 1}$$

for any $0 < \delta \leq \delta_0$ and all $n \in \mathbb{N}$ sufficiently big.

Until now we have used only Lebesgue measure. The next (easy) lemma allows us to get the estimates for any other measure absolutely continuous with respect to the Lebesgue measure.

Lemma 5. *Let $f \in L^1(\mathbb{R}^n)$ and let (K_r) , $r > 0$, be a system of measurable sets $K_r \subset \mathbb{R}^n$ satisfying the following condition:*

$$\text{There is } \sigma > 0 \text{ such that for every cube } Q \subset \mathbb{R}^n, \limsup_{r \rightarrow 0} \frac{\mathcal{L}_n(Q \cap K_r)}{\mathcal{L}_n Q} \leq \sigma.$$

Then

$$\limsup_{r \rightarrow 0} \int_{K_r} f \, d\mathcal{L}_n \leq \sigma \|f\|_{L^1}.$$

Proof. Let $\varepsilon > 0$. There is a continuous function $g: \mathbb{R}^n \rightarrow \mathbb{R}$ with compact support such that $\|f - g\|_{L^1} \leq \varepsilon$. Further, by the uniform continuity of g , there is $\delta > 0$ such that

$$|g(x) - g(y)| \leq \varepsilon$$

for any $x, y \in \mathbb{R}^n$ satisfying $\|x - y\| \leq \delta$.

Let $Q \subset \mathbb{R}^n$ be a cube containing the support of g . We partition the cube Q into a finite family \mathcal{P} of subcubes of diameter at most δ , and then we choose in each

$P \in \mathcal{P}$ a point $x_P \in P$, for example the centre. Now

$$\begin{aligned}
 \int_{K_r} f d\mathcal{L}_n &\leq \|f - g\|_{L^1} + \int_{K_r} g d\mathcal{L}_n \leq \varepsilon + \sum_{P \in \mathcal{P}} \int_{P \cap K_r} g d\mathcal{L}_n \\
 &\leq \varepsilon + \sum_{P \in \mathcal{P}} \int_{P \cap K_r} (g - g(x_P)) d\mathcal{L}_n + \sum_{P \in \mathcal{P}} g(x_P) \mathcal{L}_n(P \cap K_r) \\
 &\leq \varepsilon + \varepsilon \sum_{P \in \mathcal{P}} \mathcal{L}_n(P \cap K_r) + \sum_{P \in \mathcal{P}} g(x_P) \mathcal{L}_n(P \cap K_r) \\
 (15) \quad &\leq \varepsilon + \varepsilon \mathcal{L}_n Q + \sum_{P \in \mathcal{P}} g(x_P) \mathcal{L}_n(P \cap K_r).
 \end{aligned}$$

By the assumption we can choose $r > 0$ small enough to guarantee that

$$\mathcal{L}_n(P \cap K_r) \leq (\sigma + \varepsilon) \mathcal{L}_n P$$

for all $P \in \mathcal{P}$. Then the last sum in (15) can be estimated by

$$\begin{aligned}
 \sum_{P \in \mathcal{P}} g(x_P) \mathcal{L}_n(P \cap K_r) &\leq (\sigma + \varepsilon) \sum_{P \in \mathcal{P}} g(x_P) \mathcal{L}_n P \\
 &\leq (\sigma + \varepsilon) \left(\int_Q g d\mathcal{L}_n + \sum_{P \in \mathcal{P}} \int_P (g(x_P) - g) d\mathcal{L}_n \right) \\
 &\leq (\sigma + \varepsilon) \left(\int_Q g d\mathcal{L}_n + \varepsilon \mathcal{L}_n Q \right) \\
 &\leq (\sigma + \varepsilon) \left(\|f - g\|_{L^1} + \|f\|_{L^1} + \varepsilon \mathcal{L}_n Q \right) \\
 &\leq (\sigma + \varepsilon) \left(\varepsilon + \|f\|_{L^1} + \varepsilon \mathcal{L}_n Q \right).
 \end{aligned}$$

Combining this estimate with the estimate in (15), we obtain that

$$\limsup_{r \rightarrow 0} \int_{K_r} f d\mathcal{L}_n \leq \varepsilon + \varepsilon \mathcal{L}_n Q + (\sigma + \varepsilon) \left(\varepsilon + \|f\|_{L^1} + \varepsilon \mathcal{L}_n Q \right).$$

Since $\varepsilon > 0$ is arbitrarily small we conclude that

$$\limsup_{r \rightarrow 0} \int_{K_r} f d\mathcal{L}_n \leq \sigma \|f\|_{L^1},$$

which completes the proof. \square

Proof of Proposition 1. Let $\varepsilon > 0$ be given. We choose $\delta \in (0, \delta_0]$ such that the conclusion of Lemma 4 holds true. Also, let $n \in \mathbb{N}$ be large enough to satisfy

$$(16) \quad \frac{1}{\frac{1}{2}(1 + \delta)^n - 1} \leq \frac{\varepsilon}{3}.$$

By assumption, there is a finite-dimensional space U , $\dim U \geq n$, such that μ_U is absolutely continuous with respect to the Lebesgue measure \mathcal{L}_U . We denote

$$f = \frac{d\mu_U}{d\mathcal{L}_U}.$$

For every $r > 0$ there is a $\mathfrak{B}^{(r)} \in \mathfrak{B}(r)$ such that

$$(17) \quad \mu_U \left(\bigcup \mathfrak{B}_U^{(r)} \right) \geq \frac{1}{2} \sup \left\{ \mu_U \left(\bigcup \mathfrak{B}_U \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\}.$$

We put $K_r = \bigcup \mathfrak{B}_U^{(r)}$, $r > 0$. For this choice of K_r the assumption of Lemma 5 is satisfied: Let $Q \subset U$ be a cube. Then by Lemma 4 in the form (14) and by (16) we have

$$\limsup_{r \rightarrow 0} \frac{\mathcal{L}_U(Q \cap \bigcup \mathfrak{B}_U^{(r)})}{\mathcal{L}_U Q} \leq \limsup_{r \rightarrow 0} \sup_{\mathfrak{B} \in \mathfrak{B}(r)} \frac{\mathcal{L}_U(Q \cap \bigcup \mathfrak{B}_U)}{\mathcal{L}_U Q} \leq \frac{\varepsilon}{3}.$$

So Lemma 5 implies that

$$\limsup_{r \rightarrow 0} \mu_U \left(\bigcup \mathfrak{B}_U^{(r)} \right) = \limsup_{r \rightarrow 0} \int_{\bigcup \mathfrak{B}_U^{(r)}} f d\mathcal{L}_U \leq \frac{\varepsilon}{3}.$$

In combination with (17) we get

$$\limsup_{r \rightarrow 0} \sup \left\{ \mu_U \left(\bigcup \mathfrak{B}_U \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \leq \frac{2\varepsilon}{3}.$$

It follows that there is $r_0 > 0$ such that for all $0 < r \leq r_0$ we have

$$(18) \quad \sup \left\{ \mu_U \left(\bigcup \mathfrak{B}_U \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \leq \varepsilon.$$

Now we are ready to estimate the measure $\mu \left(\bigcup \mathfrak{B} \right)$ for any $\mathfrak{B} \in \mathfrak{B}(r)$ and $0 < r \leq r_0$. By condition (iii) in Proposition 1 and (18),

$$\begin{aligned} \mu \left(\bigcup \mathfrak{B} \right) &\leq (\mu_U \times \mu_{U^\perp}) \left(\bigcup \mathfrak{B} \right) = \int_{U^\perp} \mu_U \left(U \cap (x + \bigcup \mathfrak{B}) \right) d\mu_{U^\perp}(x) \\ &\leq \int_{U^\perp} \sup \left\{ \mu_U \left(\bigcup \mathfrak{B}_U \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} d\mu_{U^\perp} \leq \int_{U^\perp} \varepsilon d\mu_{U^\perp} = \varepsilon. \end{aligned}$$

Since this is true for all $\mathfrak{B} \in \mathfrak{B}(r)$ we may conclude that

$$\sup \left\{ \mu \left(\bigcup \mathfrak{B} \right) \mid \mathfrak{B} \in \mathfrak{B}(r) \right\} \leq \varepsilon,$$

provided that $0 < r \leq r_0$ and Proposition 1 is proved. \square

It may be of some interest to make the following final remark. Although the classical version of the Vitali Covering Theorem fails for e.g. all infinite-dimensional Gaussian measures, there is still a weaker statement of the covering type which holds true. The validity of the Differentiation theorem is, in fact, equivalent to such a weak covering theorem. For details see e.g. Hayes and Pauc [2], or the deep paper of M. Talagrand [7], where this connection is treated in considerable generality.

Based on the already-mentioned positive differentiation result [8] for some class \mathcal{G} of Gaussian measures we have the following:

Given $\gamma \in \mathcal{G}$, $1 < p < \infty$, $\varepsilon > 0$, and Vitali system \mathcal{V} on a set A in a separable Hilbert space, there is a countable subsystem $\mathcal{S} \subset \mathcal{V}$ such that

- (i) $\gamma(A \setminus \bigcup \mathcal{S}) = 0$,
- (ii) $\left\| \sum_{B \in \mathcal{S}} \chi_B - \chi_{\bigcup \mathcal{S}} \right\|_{L^p(\gamma)} < \varepsilon$.

Condition (ii) means that instead of disjointness we are only able to make the overlap of sets in \mathcal{S} arbitrarily small in the given L^p norm.

REFERENCES

- [1] de Guzmán, M., *Real Variable Methods in Fourier Analysis*, Notas de Matemática (75), North-Holland Publishing Co., Amsterdam, 1981. MR **83j**:42019
- [2] Hayes, C.A. and Pauc, C.Y., *Derivation and Martingales*, Springer-Verlag, 1970
- [3] Kuo, H.-H., *Gaussian Measures in Banach Spaces*, Lecture Notes in Math. 463, Springer-Verlag, 1975 MR **57**:1628
- [4] Preiss, D., *Gaussian measures and covering theorems*, Comment. Math. Univ. Carolinae 20(1979), 95–99 MR **80e**:28020
- [5] Preiss, D., *Gaussian measures and the density theorem*, Comment. Math. Univ. Carolinae 22(1981), 181–193 MR **83d**:28008
- [6] Preiss, D., *Differentiation of measures in infinitely-dimensional spaces*, Proc. Topology and Measure III (Greifswald, 1982), Wissen. Beiträge d. Greifswald Univ., 1983, 201–207 MR **85a**:60015
- [7] Talagrand, M., *Derivation, L^Ψ -bounded martingales and covering conditions*, Trans. Amer. Math. Soc. 293, no. 1(1986), 257–291 MR **87c**:28005
- [8] Tišer, J., *Differentiation theorem for Gaussian measures on Hilbert space*, Trans. Amer. Math. Soc. 308, no. 2(1988), 655–666 MR **89m**:28023
- [9] Vitali, G., *Sui gruppi di punti a sulle funzioni di variabili reali*, Atti Accad. Sci. Torino 43 (1908), 229–246

DEPARTMENT OF MATHEMATICS, FACULTY OF ELECTRICAL ENGINEERING, CZECH TECHNICAL UNIVERSITY, TECHNICKÁ 2, 166 27 PRAHA, CZECH REPUBLIC

E-mail address: tiser@math.feld.cvut.cz

ON IDEMPOTENTS IN REDUCED ENVELOPING ALGEBRAS

GEORGE B. SELIGMAN

ABSTRACT. Explicit constructions are given for idempotents that generate all projective indecomposable modules for certain finite-dimensional quotients of the universal enveloping algebra of the Lie algebra $sl(2)$ in odd prime characteristic. The program is put in a general context, although constructions are only carried through in the case of $sl(2)$.

§1. INTRODUCTION

We consider the Lie algebra $sl(2)$ with basis e, f, h , relations $[ef] = h$, $[eh] = 2e$, $[fh] = -2f$. A reduced enveloping algebra for the Lie algebra $\mathfrak{g} = sl(2)$ over a field F , assumed algebraically closed and of odd prime characteristic p , is determined by three parameters $\lambda, \mu, \nu \in F$. It is the algebra \mathfrak{A} defined as the quotient of the universal enveloping algebra $\mathfrak{U} = \mathfrak{U}(sl(2))$ by the ideal generated by the (central) elements $h^p - h - \lambda$, $e^p - \mu$, $f^p - \nu$ of \mathfrak{U} . The dimension of \mathfrak{A} is p^3 . We identify e, f, h with their images in \mathfrak{A} whenever there is no great danger of confusion.

The subalgebra $F[h]$ of \mathfrak{A} is semisimple; since the minimum polynomial $X^p - X - \lambda$ of h factors as $\prod_{\alpha \in \mathbf{F}_p} (X - \tau - \alpha)$, where $\tau \in F$ is fixed with $\tau^p - \tau = \lambda$ (we fix $\tau = 0$ if $\lambda = 0$), we let $g_\alpha(X) = \frac{X^p - X - \lambda}{X - \tau - \alpha}$ for each $\alpha \in \mathbf{F}_p$. Then $h(\alpha) = g_\alpha(\tau + \alpha)^{-1} g_\alpha(h)$ is a nonzero idempotent in $F[h]$, and these $h(\alpha)$ form a complete system of orthogonal idempotents in $F[h]$: $h(\alpha)h(\beta) = 0$ if $\alpha \neq \beta$, $\sum_{\alpha \in \mathbf{F}_p} h(\alpha) = 1$, $h(\alpha)h = (\tau + \alpha)h(\alpha)$. A

generalization in the case of the restricted enveloping algebra, yielding a complete set of idempotents in the enveloping algebra of a Cartan subalgebra, was given by Nielsen [N] in 1963. In that thesis, Nielsen also gives generators for minimal one-sided ideals in the restricted enveloping algebra of $\mathfrak{sl}(2)$.

We propose to refine this set of idempotents to a complete set of primitive idempotents in \mathfrak{A} . We work in the commutative subalgebra of \mathfrak{A} generated by h and ef , and in the commutative subalgebra \mathfrak{W} of \mathfrak{U} generated by h, ef, e^p and f^p . (The subalgebra \mathfrak{W} will only be involved in §5, where we shall recall that certain elements of \mathfrak{W} satisfy relations, such as belonging to the ideal \mathfrak{J} in \mathfrak{W} generated by $e^p - \mu, f^p - \nu, h^p - h - \lambda$. For example, the elements $h^*(\alpha) = g_\alpha(\tau + \alpha)^{-1} g_\alpha(h)$, now regarded as elements of \mathfrak{W} , still satisfy $\sum_{\alpha \in \mathbf{F}_p} h^*(\alpha) = 1 \in \mathfrak{W}$, with $h^*(\alpha)h^*(\beta) - \delta_{\alpha\beta}h^*(\beta)$ in the ideal in \mathfrak{W} generated by $h^p - h - \lambda$, as is $h^*(\alpha)h - (\tau + \alpha)h^*(\alpha)$.)

Received by the editors August 14, 2002 and, in revised form, January 15, 2003.

2000 *Mathematics Subject Classification*. Primary 17B35, 16S30.

The key to the construction is the following observation, easily proved by induction:

Lemma 1. *For any set of i or more indeterminates X_1, \dots, X_k , let $s_i[X_1, \dots, X_k]$ be the i -th elementary symmetric function ($s_0 = 1$). Then for every positive integer j , and each $\alpha \in \mathbf{F}_p$,*

$$(1) \quad h(\alpha)e^j f^j = \sum_{k=0}^{j-1} s_k(\alpha + \tau + 2, 2(\alpha + \tau + 3), \dots, (j-1)(\alpha + \tau + j))(h(\alpha)ef)^{j-k}.$$

We indicate a proof, with $h(\alpha)$ replaced by $h^*(\alpha)$, and working in \mathfrak{U} , that both sides of the relation are in \mathfrak{W} and are congruent modulo the ideal in \mathfrak{W} as above. The relation in \mathfrak{A} then follows. For $j = 1$, there is nothing to prove. If one writes

$$h^*(\alpha)e^{j+1}f^{j+1} = h^*(\alpha)e^j f^j ef + \sum_{k=0}^j h^*(\alpha)e^j f^k [ef] f^{j-k}$$

and uses

$$[ef] = h, h^*(\alpha)e^j f^k h f^{j-k} = h^*(\alpha)h e^j f^j + 2(j-k)h^*(\alpha)e^j f^j,$$

one deduces inductively that all $h^*(\alpha)e^m f^m$ are in \mathfrak{W} . Furthermore, all $e^m f^m$ are in \mathfrak{W} , so that $h^*(\alpha)h e^j f^j$ is congruent, modulo our ideal \mathfrak{J} in \mathfrak{W} , to $(\tau + \alpha)h^*(\alpha)e^j f^j$. Thus

$$h^*(\alpha)e^{j+1}f^{j+1} \equiv h^*(\alpha)e^j f^j ef + j(\alpha + \tau + j + 1)h^*(\alpha)e^j f^j \pmod{\mathfrak{J}},$$

giving the necessary inductive step. When $j = p$ (still working in \mathfrak{W}), we have $h^*(\alpha)e^p f^p \equiv \mu\nu h^*(\alpha) \pmod{\mathfrak{J}}$. If $b \in h^*(\alpha)\mathfrak{W}$, then $h^*(\alpha)b \equiv b \pmod{\mathfrak{J}}$, so that

$$\mu\nu h^*(\alpha) \equiv \sum_{k=0}^{p-1} s_k(\tau + \alpha + 2, 2(\tau + \alpha + 3), \dots, (p-1)(\tau + \alpha + p))(h^*(\alpha)ef)^{p-k} \pmod{\mathfrak{J}}.$$

In the homomorphic image in \mathfrak{A} of \mathfrak{W} , the image $h(\alpha)$ of $h^*(\alpha)$ is the unit element for the ideal it generates, and the element $h(\alpha)ef$ of this ideal satisfies the polynomial equation $f(X) = 0$ relative to the unit element $h(\alpha)$, where

$$f(X) = \prod_{\alpha \in \mathbf{F}_p} (X + i(\tau + \alpha + i + 1)) - \mu\nu.$$

Changing the variable here to $Y = X - (\frac{\tau + \alpha + 1}{2})^2$, we have

$$f(X) = g(Y) = \prod_{i \in \mathbf{F}_p} (Y + ((\frac{\tau + \alpha + 1}{2}) + i)^2) - \mu\nu.$$

Now it is a straightforward exercise to verify that, for indeterminates Y and Z over \mathbf{F}_p ,

$$\prod_{i \in \mathbf{F}_p} (Y + (i + Z)^2) = Y(Y^{\frac{p-1}{2}} + (-1)^{\frac{p+1}{2}})^2 + (Z^p - Z)^2.$$

Thus

$$\begin{aligned} (2) \quad g(Y) &= Y(Y^{\frac{p-1}{2}} + (-1)^{\frac{p+1}{2}})^2 + (\frac{\tau^p - \tau}{2})^2 - \mu\nu \\ &= Y(Y^{\frac{p-1}{2}} + (-1)^{\frac{p+1}{2}})^2 + (\frac{\lambda^2}{4} - \mu\nu). \end{aligned}$$

Here $g'(Y) = (-1)^{\frac{p+1}{2}} Y^{\frac{p-1}{2}} + 1$, so $g(Y)$ is semisimple if and only if no element $-\beta^2$ ($0 \neq \beta \in \mathbf{F}_p$) is a root of $g(Y)$. One easily checks that

$$g(-\beta^2) = -\beta^2 + 2\beta^2 - \beta^2 + \frac{1}{4}\lambda^2 - \mu\nu = \frac{1}{4}\lambda^2 - \mu\nu.$$

That is, $g(Y)$ has repeated roots if and only if $\lambda^2 = 4\mu\nu$. We refer to $\lambda^2 - 4\mu\nu$ as the discriminant of \mathfrak{A} .

§2. THE SEMISIMPLE CASE

When the discriminant is nonzero, $f(X)$ has p distinct roots $\theta_1, \dots, \theta_p$. Lagrange interpolation may be applied once more to obtain $h(\alpha)$ (or $h^*(\alpha)$, modulo \mathfrak{J}) as a sum of p elements in $F[h(\alpha), h(\alpha)ef]$ that are nonzero orthogonal idempotents (with the same holding with $h(\alpha)$ replaced by $h^*(\alpha)$, if we work in \mathfrak{W} , modulo \mathfrak{J}). As α runs over \mathbf{F}_p , one obtains p^2 orthogonal idempotents in \mathfrak{A} , with sum 1. Typical for these is an element $E(\alpha, \theta_i) = h(\alpha)q(h(\alpha)ef)$, where $q(X)$ is a polynomial of degree $p-1$, depending on α and θ_i , with $E(\alpha, \theta_i)h(\alpha)ef = \theta_i E(\alpha, \theta_i)$ in \mathfrak{A} (or with a corresponding congruence in \mathfrak{W} , modulo \mathfrak{J}).

Under these circumstances, it is easy to see that every irreducible \mathfrak{A} -module \mathfrak{M} has dimension at least (indeed, equal to) p : If one of μ, ν is nonzero and if v is an eigenvector for h in \mathfrak{M} , then so are all ve^k, vf^k , and one of these families yields eigenvectors for p distinct eigenvalues. If $\mu = 0 = \nu, vh = \kappa v$, we may assume $ve = 0$. Then the usual treatment of $sl(2)$ -modules shows that if the first integer $k > 0$ with $vf^k = 0$ is less than p , then $\kappa \in \mathbf{F}_p$ and $\lambda = 0$, a contradiction. Thus all $vf^k, 0 \leq k \leq p$, are linearly independent. (More detail on irreducible modules will be needed below; for a more thorough and extensive study of irreducible modules in all reductive cases, see [JCJ].)

It follows that the right ideals $E(\alpha, \theta_i)\mathfrak{A}$ have dimension at least p . But there are p^2 such ideals, and the algebra \mathfrak{A} , of dimension p^3 , is their direct sum. Thus each $E(\alpha, \theta_i)\mathfrak{A}$ is a minimal right ideal in \mathfrak{A} , and \mathfrak{A} is a semisimple algebra. From general principles, the modules $E(\alpha, \theta_i)\mathfrak{A}$ group into p blocks, each consisting of p isomorphic \mathfrak{A} -modules. We determine conditions for isomorphism in terms of the parameters α and θ_i .

First, the promised details on irreducible \mathfrak{A} -modules: If \mathfrak{M} is such a module, then h and ef commute in their actions, so have a common eigenvector v . As above, if $\mu \neq 0$ or $\nu \neq 0$, then all ve^j or all $vf^j, 0 \leq j < p$ are linearly independent. One readily verifies that their span is stable under the action of e, f, h , so is equal to \mathfrak{M} . If $\mu = \nu = 0$, then $ve^r \neq 0, ve^{r+1} = 0$ for some $r, 0 \leq r < p$. Let $w = ve^r$. Then w, wf, \dots, wf^{p-1} form a basis for \mathfrak{M} . Having fixed τ (necessarily an eigenvalue of h), the eigenvalue of ef to which a τ -eigenvector for h belongs determines \mathfrak{M} up to isomorphism.

In the present case, $E(\alpha, \theta)ef = E(\alpha, \theta)h(\alpha)ef = \theta E(\alpha, \theta)$, and there is some e^j or f^k (possibly both) with $E(\alpha, \theta)e^j$ or $E(\alpha, \theta)f^k$ an h -eigenvector of eigenvalue τ . The commutation relations then yield that the corresponding eigenvalue of ef is $\theta - \frac{\alpha}{2}(\tau + \frac{\alpha}{2} + 1)$. (Independently of whether we use $E(\alpha, \theta)ef$ or $E(\alpha, \theta)f^k$; for instance, if $0 \neq E(\alpha, \theta)f^k$ is an h -eigenvector belonging to the h -eigenvalue τ , then $2k = \alpha$ [in \mathbf{F}_p], and, inductively on m , $E(\alpha, \theta)f^m ef = (\theta - m\tau - m\alpha + m(m-1))E(\alpha, \theta)f^m$.)

Accordingly, we have $E(\alpha, \theta)\mathfrak{A} \cong E(\beta, \theta')\mathfrak{A}$ if and only if

$$(3) \quad \theta - \frac{\alpha}{2}(\tau + \frac{\alpha}{2} + 1) = \theta' - \frac{\beta}{2}(\tau + \frac{\beta}{2} + 1).$$

It may help to clarify this statement by noting that if θ is a root of $f(X) = \prod_{i \in \mathbf{F}_p} (X + i(\tau + \alpha + i + 1)) - \mu\nu$ as before, and if we make the change of variable

$$X = X' + \frac{\alpha}{2}(\tau + \frac{\alpha}{2} + 1) - \frac{\beta}{2}(\tau + \frac{\beta}{2} + 1),$$

then

$$\begin{aligned} f(X) &= f^*(X') = \prod_{i \in \mathbf{F}_p} (X' + (i + \frac{\alpha}{2} - \frac{\beta}{2})(\tau + \beta + (i + \frac{\alpha}{2} - \frac{\beta}{2}) + 1)) - \mu\nu \\ &= \prod_{j \in \mathbf{F}_p} (X' + j(\tau + \beta + j + 1)) - \mu\nu, \end{aligned}$$

so that $\theta' = \theta - \frac{\alpha}{2}(\tau + \frac{\alpha}{2} + 1) + \frac{\beta}{2}(\tau + \frac{\beta}{2} + 1)$ is a root of $f^*(X')$. For each $\beta \neq \alpha$ in \mathbf{F}_p , there is one such θ' . Thus, for given α , all $E(\alpha, \theta)\mathfrak{A}$ are non-isomorphic \mathfrak{A} -modules, while for each pair α, β ($\alpha \neq \beta$), the isomorphism classes of modules $E(\alpha, \theta)\mathfrak{A}$ and $E(\beta, \theta')\mathfrak{A}$ are the same when the relation (3) holds.

§3. THE DEGENERATE, NON-RESTRICTED CASE

When $\lambda^2 = 4\mu\nu$,

$$f(X) = (X - (\frac{\tau + \alpha + 1}{2})^2)((X - (\frac{\tau + \alpha + 1}{2})^2)^{\frac{p-1}{2}} + (-1)^{\frac{p+1}{2}})^2$$

(by (2)) is the minimum polynomial of $h(\alpha)ef$ relative to the unit element $h(\alpha)$. The roots in F of $Y^{\frac{p-1}{2}} + (-1)^{\frac{p-1}{2}} = 0$ in F are the negatives of nonzero squares in \mathbf{F}_p . So

$$f(X) = (X - (\frac{\tau + \alpha + 1}{2})^2) \prod_{\gamma \in \mathbf{F}_p^{*2}} (X - (\frac{\tau + \alpha + 1}{2})^2 + \gamma)^2.$$

The following lemma, whose proof is left as an exercise, gives a resolution of each $h(\alpha)$ into orthogonal idempotents:

Lemma 2. Let $g_0(Y) = \prod_{\gamma \in \mathbf{F}_p^{*2}} (Y + \gamma)^2$, and, for $\beta \in \mathbf{F}_p^{*2}$, let $g_\beta(Y) = 2Y(Y - \beta) \prod_{\substack{\gamma \in \mathbf{F}_p^{*2} \\ \gamma \neq \beta}} (Y + \gamma)^2$. Then

$$(4) \quad g_0(Y) + \sum_{\beta \in \mathbf{F}_p^{*2}} g_\beta(Y) = 1, \text{ and}$$

$$g_\xi(Y)g_\eta(Y) \equiv \delta_{\xi\eta}g_\eta(Y) \pmod{f(Y + (\frac{\tau + \alpha + 1}{2})^2)} = g(Y).$$

It follows that the elements $E(\alpha, 0) = f_0(h(\alpha)ef)$, $E(\alpha, \beta) = f_\beta(h(\alpha)ef)$, where $f_\gamma(X) = g_\gamma(X - (\frac{\tau + \alpha + 1}{2})^2)$, form a set of $\frac{p+1}{2}$ (nonzero) orthogonal idempotents in $F[h(\alpha), h(\alpha)ef] \subset \mathfrak{A}$, whose sum is the unit element $h(\alpha)$. (When $h(\alpha)$ is replaced by $h^*(\alpha)$, the corresponding relations hold (mod \mathfrak{J}) in the subalgebra $f[h^*(\alpha), h^*(\alpha)ef]$ of the commutative subalgebra \mathfrak{W} of $\mathfrak{U}(\mathfrak{g})$.) As α runs over \mathbf{F}_p ,

we obtain a family of $\frac{p(p+1)}{2}$ orthogonal idempotents $E(\alpha, \beta)$ in $F[h, ef] \subset \mathfrak{A}$, whose sum is 1.

To see that these idempotents are *primitive*, one may reason as follows: Each $E(\alpha, 0)$, $\alpha \in \mathbf{F}_p$, generates a nonzero right \mathfrak{A} -module. If we assume not all of λ, μ, ν are 0, then as before, this module has dimension at least p , and dimension exactly p if and only if it is irreducible. We defer the "restricted" case $\lambda = \mu = \nu = 0$ to the next section. When $\beta \in \mathbf{F}_p^{*2}$, left-multiplication by $h(\alpha)ef - (\frac{\tau+\alpha+1}{2})^2 h(\alpha) + \beta h(\alpha)$ is a nonzero nilpotent \mathfrak{A} -endomorphism of $E(\alpha, \beta)\mathfrak{A}$ whose image lies in its kernel, and where each of the kernel and cokernel has dimension at least p . Thus $E(\alpha, \beta)\mathfrak{A}$ has dimension at least $2p$, and the dimension is exactly $2p$ only if both the image and kernel are irreducible. In that case the endomorphism above induces an isomorphism of the cokernel onto the kernel. The sum of the dimensions of all $E(\alpha, \beta)\mathfrak{A}$ is thus at least $p^2 + p \cdot \frac{p-1}{2} \cdot 2p = p^3 = \dim \mathfrak{A}$, with equality if and only if each $E(\alpha, 0)\mathfrak{A}$ is irreducible and each $E(\alpha, \beta)\mathfrak{A}$, $\beta \in \mathbf{F}_p^{*2}$ has dimension $2p$. From the above, the latter modules are indecomposable. We have the

Theorem 1. *If $\lambda^2 = 4\mu\nu$, not all of λ, μ, ν being equal to zero, fix a solution τ of $X^p - X - \lambda = 0$, with $\tau = 0$ if $\lambda = 0$. Form elements $E(\alpha, \eta)$ for all $\eta \in \mathbf{F}_p^2$, $\alpha \in \mathbf{F}_p$, as above. Then these $\frac{p(p+1)}{2}$ elements of $F[h, ef] \subset \mathfrak{A}$ form a complete set of primitive idempotents in \mathfrak{A} . The right ideals $E(\alpha, 0)\mathfrak{A}$ are isomorphic irreducible \mathfrak{A} -modules. The right ideals $E(\alpha, \beta)\mathfrak{A}$, $\beta \in \mathbf{F}_p^{*2}$, are (projective) indecomposable \mathfrak{A} -modules of length 2, each with isomorphic composition factors. Two modules $E(\alpha, \eta)\mathfrak{A}$ and $E(\alpha', \eta')\mathfrak{A}$ are isomorphic if and only if $\eta = \eta'$.*

Proof. Only the assertions about isomorphisms remain to be proved. Here we may assume $\mu \neq 0$; the argument for $\nu \neq 0$ is analogous, and both cannot be zero because $\lambda^2 = 4\mu\nu$. (Some isomorphisms, e.g., those of all $E(\alpha, 0)\mathfrak{A}$, follow from general theory, but our identification gives a little more detail.)

An element of the irreducible module $E(\alpha, 0)\mathfrak{A}$ of h -eigenvalue τ is $E(\alpha, 0)e^j$, where $2j$ represents $-\alpha \pmod{p}$, and

$$E(\alpha, 0)e^j ef = ((\frac{\tau + \alpha + 1}{2})^2 + j(\tau + \alpha + j + 1))E(\alpha, 0)e^j.$$

But $j(\tau + \alpha + j + 1) = -\frac{\alpha}{2}(\tau + \alpha - \frac{\alpha}{2} + 1)$, so that $(\frac{\tau + \alpha + 1}{2})^2 + j(\tau + \alpha + j + 1) = (\frac{\tau + 1}{2})^2$, independent of α . Thus all $E(\alpha, 0)\mathfrak{A}$ are isomorphic.

Similarly, the quotient of $E(\alpha, \beta)\mathfrak{A}$, for $\beta \in \mathbf{F}_p^{*2}$, by its unique maximal submodule has as h -eigenvector of eigenvalue τ the coset of $E(\alpha, \beta)e^j$, $2j = -\alpha$ as above, and

$$E(\alpha, \beta)e^j ef \equiv (-\beta + (\frac{\tau + 1}{2})^2)E(\alpha, \beta)e^j,$$

as before, here modulo the maximal submodule. By [C-R], Theorem 54.11, if \mathfrak{N} is the radical of \mathfrak{A} , $E(\alpha, \beta)\mathfrak{N}$ is the unique maximal submodule of $E(\alpha, \beta)\mathfrak{A}$, and $E(\alpha, \beta)\mathfrak{A}$ and $E(\alpha', \beta')\mathfrak{A}$ are isomorphic if and only if

$$E(\alpha, \beta)\mathfrak{A}/E(\alpha, \beta)\mathfrak{N} \cong E(\alpha', \beta')\mathfrak{A}/E(\alpha', \beta')\mathfrak{N}.$$

We have just seen that this last isomorphism holds if and only if $\beta = \beta'$. This completes the proof. \square

§4. THE RESTRICTED CASE

Now let $\lambda = \mu = \nu = 0$. In this case, the projective indecomposable \mathfrak{A} -modules are more complicated, but their structure is well known (see [P]). There is the irreducible Steinberg module, generated by an element v with $ve = 0, vh = -v$, and basis v, vf, \dots, vf^{p-1} ; and there are $p-1$ modules, each of dimension $2p$ and length 4. We assign to the Steinberg module, " \mathfrak{M}_{p-1} ", the parameter $p-1$, and parameters $0, \dots, p-2$ to the remaining modules, as follows:

The module \mathfrak{M}_k has generator u_k , with $u_k h = k u_k$. To give further relations, and for future reference, we define, for $\alpha \in \mathbf{F}_p$, $\text{res}(\alpha)$ to be the ordinary integer $j, 0 \leq j < p$, whose residue class (mod p) is α . Now let $k' = p-2-k$ ($\neq k \pmod{p}$) and let $r_k = \text{res}(\frac{k'-k}{2})$. Let $v_{k'} = u_k e^{r_k}$.

Then $v_{k'} \neq 0$ but $v_{k'} e = 0$, and

$$v_{k'} f^{r_k-1} = (-1)^{r_k-1} (r_k-1)! u_k e.$$

A basis for \mathfrak{M}_k consists of the elements $\{u_k f^i, v_{k'} f^i\}, 0 \leq i < p$. The unique maximal submodule of \mathfrak{M}_k is $\langle v_{k'} \rangle + \langle u_k f^{k+1} \rangle$, of codimension $k+1$.

Here $\tau = 0$, and again, we have $\frac{p(p+1)}{2}$ orthogonal idempotents $E(\alpha, \beta)$ in \mathfrak{A} , given as in §3. The algebra \mathfrak{A} is a Frobenius algebra [Ber], indeed a symmetric algebra [S]. So the multiplicity of \mathfrak{M}_k as an indecomposable summand of \mathfrak{A} is equal to the dimension of the quotient of \mathfrak{M}_k by its maximal submodule, or $k+1$ ([C-R], Theorem 61.13). It follows that the number of indecomposable summands of \mathfrak{A} is

$$p + \sum_{k=0}^{p-2} (k+1) = \frac{p(p+1)}{2},$$

and therefore that all our (nonzero) $\frac{p(p+1)}{2}$ orthogonal idempotents are primitive, and all $E(\alpha, \beta)\mathfrak{A}$ are indecomposable.

Theorem 2. *When $\lambda = \mu = \nu = 0$, all $E(\alpha, 0)\mathfrak{A}$ are isomorphic for $\alpha \in \mathbf{F}_p$, and are the (irreducible) Steinberg module \mathfrak{M}_{p-1} . For $1 \leq j \leq \frac{p-1}{2}$, the modules $E(\alpha, j^2)\mathfrak{A}$ are indecomposable, with $E(\alpha, j^2)\mathfrak{A}$ isomorphic to $\mathfrak{M}_{\text{res}(-2j-1)}$ or \mathfrak{M}_{2j-1} , according as $j \leq \text{res}(-\frac{\alpha+1}{2}) < p-j$, or not. The $E(\alpha, \beta), \alpha \in \mathbf{F}_p, \beta \in \mathbf{F}_p^2$, are a complete set of primitive idempotents in \mathfrak{A} .*

Proof. The last assertion has been established. If r is minimal with $E(\alpha, 0)e^{r+1} = 0$, then

$$0 = E(\alpha, 0)e^{r+1}f = \left(\frac{\alpha + 2r + 1}{2}\right)^2 E(\alpha, 0)e^r.$$

So $\alpha \equiv -2r-1 \pmod{p}$, and $w = E(\alpha, 0)e^r$ has $wh = -w, we = 0$. It follows that w, wf, \dots, wf^{p-1} form a basis for an irreducible submodule isomorphic to \mathfrak{M}_{p-1} , with

$$wf^r = E(\alpha, 0)e^r f^r = (r!)^2 E(\alpha, 0) \neq 0.$$

Thus $E(\alpha, 0)\mathfrak{A} = \langle w \rangle \cong \mathfrak{M}_{p-1}$. □

For $1 \leq j \leq \frac{p-1}{2}$, we know that $E(\alpha, j^2)\mathfrak{A} \cong \mathfrak{M}_k$ for some $k, 0 \leq k \leq p-2$, and that in any such isomorphism the element $E(\alpha, j^2)$ cannot correspond to a member of the proper submodule $\langle v_{k'} \rangle + \langle u_k f^{k+1} \rangle$ of \mathfrak{M}_k . Thus $E(\alpha, j^2)$ must correspond to an element $\xi v_{k'} f^s + \eta u_k f^t$ where $\eta \neq 0, t \leq k$ and both $k'-2s$ and $k-2t$ represent $\alpha \pmod{p}$. From the last remark, we have that $k'-2s$ equal to one of $k-2t, k-2t \pm p$. If $k'-2s = k-2t$, then $k'-k$ is even, contrary to $k+k' = p-2$.

If $k' - 2s = k - 2t + p$, then $p - k - 2 - 2s = k - 2t + p$, $2t = 2k + 2s + 2$, and $t > k$, a contradiction. Thus $p - 2 - k - 2s = k' - 2s = k - 2t - p$, $2t = 2s + 2(k - p + 1) < 2s$, and $t < s$.

Applying f^{p-1-t} to our expression corresponding to $E(\alpha, j^2)$ gives

$$E(\alpha, j^2)f^{p-1-t} = \varepsilon u_k f^{p-1} \neq 0.$$

That is, $u_k f^{p-1}$ belongs to the eigenvalue $\alpha - 2(p - 1 - t)$ of h , and is annihilated by f . Its eigenvalue is evidently equal to $k - 2(p - 1) \equiv -k'$; so $\alpha + 2 + 2t = -k'$ in \mathbf{F}_p .

In other words, for the nonnegative integer m such that $E(\alpha, j^2)f^m \neq 0$, $E(\alpha, j^2)f^{m+1} = 0$ determines k' by $k' = \text{res}(2m - \alpha)$, and thereby the isomorphism class of $E(\alpha, j^2)\mathfrak{A}$: $E(\alpha, j^2)\mathfrak{A} \cong \mathfrak{M}_k$, where $k = \text{res}(\alpha - 2m - 2)$. To determine this "m", we invoke the following:

Lemma 3. *Let*

$$g_{j^2}(X) = 2(X - j^2)X \prod_{\substack{i=1 \\ i \neq j}}^{\frac{p-1}{2}} (X + i^2)^2, \quad 1 \leq j \leq \frac{p-1}{2}.$$

Let $n(\alpha, j)$ be the largest value of k such that $f_{\alpha, k}(X) = \prod_{i=0}^{k-1} (X + i(\alpha + i + 1))$ divides $g_{j^2}(X - (\frac{\alpha+1}{2})^2)$. Then $E(\alpha, j^2)\mathfrak{A}$ is the projective indecomposable module $\mathfrak{M}_{\text{res}(\alpha+2n(\alpha, j))}$.

Proof of Lemma 3. With $n(\alpha, j)$ as defined,

$$g_{j^2}(X - (\frac{\alpha+1}{2})^2) = h_j(X - (\frac{\alpha+1}{2})^2) f_{\alpha, n(\alpha, j)}(X),$$

where $h_j(X - (\frac{\alpha+1}{2})^2)$ is a product of factors involving only $X + k^2 - (\frac{\alpha+1}{2})^2$, $1 \leq k \leq \frac{p-1}{2}$; $X - (\frac{\alpha-1}{2})^2$; and $X - j^2 - (\frac{\alpha+1}{2})^2$. In \mathfrak{A} , we have therefore

$$\begin{aligned} E(\alpha, j^2) &= h_j(h(\alpha)ef - (\frac{\alpha+1}{2})^2 h(\alpha)) f_{\alpha, n(\alpha, j)}(h(\alpha)ef) \\ &= h_j(h(\alpha)ef - (\frac{\alpha+1}{2})^2 h(\alpha)) h(\alpha) e^{n(\alpha, j)} f^{n(\alpha, j)}, \end{aligned}$$

by Lemma 1. Now $h(\alpha)ef$ commutes with all $h(\alpha)e^k f^k$, and the proof of Lemma 1 involves showing that

$$h(\alpha)e^k f^k ef = h(\alpha)e^{k+1} f^{k+1} - k(\alpha + k + 1)h(\alpha)e^k f^k.$$

Thus

$$E(\alpha, j^2) = h(\alpha)e^{n(\alpha, j)} f^{n(\alpha, j)} (h_j(h(\alpha)ef - (\frac{\alpha+1}{2})^2 h(\alpha)))$$

and

$$\begin{aligned} &h(\alpha)e^{n(\alpha, j)} f^{n(\alpha, j)} h(\alpha)ef \\ &= h(\alpha)e^{n(\alpha, j)+1} f^{n(\alpha, j)+1} - n(\alpha, j)(\alpha + n(\alpha, j) + 1)h(\alpha)e^{n(\alpha, j)} f^{n(\alpha, j)}. \end{aligned}$$

By definition, $X + n(\alpha, j)(\alpha + n(\alpha, j) + 1)$ is not among the factors of $g_{j^2}(X - (\frac{\alpha+1}{2})^2)$. Thus the coefficient of $h(\alpha)e^{n(\alpha, j)} f^{n(\alpha, j)}$ will not be zero when $E(\alpha, j^2)$ is expanded in terms of the form $h(\alpha)e^i f^i$, while the other values of "i" that occur will all be greater than $n(\alpha, j)$. Accordingly, $E(\alpha, j^2)f^{p-n(\alpha, j)-1} \neq 0$ while $E(\alpha, j^2)f^{p-n(\alpha, j)} = 0$. The lemma is proved. \square

To complete the proof of the theorem: Note that $n(\alpha, j)$ is the first integer n with $n(\alpha + n + 1) = j^2 - (\frac{\alpha+1}{2})^2$; so $n(\alpha, j) \equiv -(\frac{\alpha+1}{2}) \pm j \pmod{p}$. If $j \leq \text{res}(-(\frac{\alpha+1}{2})) < p - j$, then $n(\alpha, j) = \text{res}(-j - \frac{\alpha+1}{2})$; otherwise, $n(\alpha, j) = \text{res}(j - \frac{\alpha+1}{2})$. The remaining assertion of the theorem follows.

Remark. For example, if $j = \frac{p-1}{2}$, only $\alpha = 0$ satisfies the inequalities $\frac{p-1}{2} \leq \text{res}(-\frac{\alpha+1}{2}) < \frac{p+1}{2}$, and $\mathfrak{M}_{\text{res}(-2j-1)} = \mathfrak{M}_0$ occurs only *once*, while $\mathfrak{M}_{2j-1} = \mathfrak{M}_{p-2}$ occurs with multiplicity $p - 1$.

§5. SOME GENERALIZATIONS

The remarks of this section apply generally. The case where the underlying Lie algebra is $\mathfrak{sl}(2)$ will be observed, from earlier constructions, to satisfy the conditions imposed on the idempotents. Application to that case enables us to recover, in somewhat more explicit form, some results of Christopher Bendel [Ben].

Here \mathfrak{g} is an arbitrary Lie algebra of prime characteristic and finite dimension over a field F , assumed algebraically closed (although further assumptions avoid this constraint). Let x_1, \dots, x_n be a basis for \mathfrak{g} . For each x_i , let z_i be a p -polynomial in x_i that is central in $\mathfrak{U}(\mathfrak{g})$, say of degree p^{m_i} . Fix scalars $\lambda_1, \dots, \lambda_n \in F$ and a nonnegative integer s . Let \mathfrak{S} be the ideal in $\mathfrak{U}(\mathfrak{g})$ generated by the (central) elements $T_i^{p^s}$, $1 \leq i \leq n$, where $T_i = z_i - \lambda_i$. Let \mathfrak{T} be the ideal generated by the T_i . Then $\mathfrak{B} = \mathfrak{U}(\mathfrak{g})/\mathfrak{S}$ and $\mathfrak{A} = \mathfrak{U}(\mathfrak{g})/\mathfrak{T}$ are finite-dimensional algebras over F , of respective dimensions $p^{ns+\sum m_i}$ and $p^{\sum m_i}$ ([J], Chapter 6).

Suppose we have found elements $e_1, \dots, e_t \in \mathfrak{U}(\mathfrak{g})$ such that:

- i) $\sum_{j=1}^t e_j = 1$ in $\mathfrak{U}(\mathfrak{g})$;
- ii) there is a commutative subalgebra \mathfrak{W} of $\mathfrak{U}(\mathfrak{g})$, containing all e_j 's and all T_i 's, such that for all j, k , $e_j e_k - \delta_{jk} e_k$ is in the ideal in \mathfrak{W} generated by the T_i .

These conditions guarantee that the homomorphic images \bar{e}_j of the e_j in \mathfrak{A} form a system of orthogonal idempotents whose sum is 1. The algebra " \mathfrak{W} " previously used guarantees that the polynomials in the $h^*(\alpha)ef$ yielding idempotents in the " \mathfrak{A} " of earlier sections satisfy i) and ii).

From i) and ii), $\sum_{j=1}^t e_j^{p^s} = 1$. If $e_j e_k - \delta_{jk} e_k = \sum_{n=1}^n w_n T_n$, $w_n \in \mathfrak{W}$, then

$$e_j^{p^s} e_k^{p^s} - \delta_{jk} e_k^{p^s} = \sum w_i^{p^s} T_i^{p^s}.$$

So the homomorphic images in \mathfrak{B} of the $e_j^{p^s}$ are a system of orthogonal idempotents whose sum is 1. We denote the image of $e_j^{p^s}$ in \mathfrak{B} by E_j . If $e_j \notin \mathfrak{T}$, then $e_j^{p^s} - e_j = \sum_{r=1}^{p^s-1} (e_j^{r+1} - e_j^r) \in \mathfrak{T}$, and so $e_j^{p^s} \notin \mathfrak{S}$ and $E_j \neq 0$ if $\bar{e}_j \neq 0$.

The central subalgebra $F[T_1, \dots, T_n]$ of $\mathfrak{U}(\mathfrak{g})$ is a polynomial algebra on the T_i as generators. All monomials $T_1^{\nu_1} \dots T_n^{\nu_n}$ with some $\nu_i \geq p^s$ are in \mathfrak{S} . We order the remaining monomials $\{M_i\}$ as follows: $M < N$ if degree $M <$ degree N , and the order is linear but otherwise arbitrary among monomials of the same degree. We label by subscripts indicating place in the order. Thus

$$1 = M_0 < M_1 < M_2 < \dots < M_{p^{ns}-1} = T_1^{p^s-1} \dots T_n^{p^s-1}.$$

Then the right ideal $E_j\mathfrak{B}$ in \mathfrak{B} has descending filtration

$$E_j\mathfrak{B} = (E_j\mathfrak{B})_1 \supset \dots \supset (E_j\mathfrak{B})_{p^{ns}} \supset 0,$$

where $(E_j\mathfrak{B})_r = \sum_{q \geq r-1} E_j\mathfrak{B}M_q$.

In the right action of \mathfrak{B} (or of $\mathfrak{U}(\mathfrak{g})$) on $E_j\mathfrak{B}$, \mathfrak{T} maps each $(E_j\mathfrak{B})_r$ into $(E_j\mathfrak{B})_{r+1}$. So successive quotients in the filtration are right \mathfrak{A} -modules. Let φ be the canonical homomorphism of \mathfrak{B} into \mathfrak{A} , ψ that of $\mathfrak{U}(\mathfrak{g})$ onto \mathfrak{B} , so that $\varphi \circ \psi$ is the canonical homomorphism of $\mathfrak{U}(\mathfrak{g})$ onto \mathfrak{A} . Then

$$(\varphi \circ \psi)(e_j^{p^s} \mathfrak{U}(\mathfrak{g})) = \varphi(E_j\mathfrak{B}) = (\varphi \circ \psi)(e_j^{p^s})\mathfrak{A} = (\varphi \circ \psi)(e_j)\mathfrak{A} = \bar{e}_j\mathfrak{A}.$$

The kernel of $\varphi|_{(E_j\mathfrak{B})}$ contains $(E_j\mathfrak{B})_2$. So φ induces a homomorphism of $(E_j\mathfrak{B})_1/(E_j\mathfrak{B})_2$ onto $\bar{e}_j\mathfrak{A}$ (as $\mathfrak{U}(\mathfrak{g})$ -modules, or as \mathfrak{A} -modules). Now $\sum_{r>0} \mathfrak{B}M_r = \text{Ker } \varphi$ (by Poincaré-Birkhoff-Witt - see [J], Chapter 5), and $E_j\mathfrak{B} \cap \sum_{r>0} \mathfrak{B}M_r = \sum_{r>0} E_j\mathfrak{B}M_r$, by left-multiplying with the idempotents E_k . Thus the kernel of the induced map of $(E_j\mathfrak{B})_1$ onto $\bar{e}_j\mathfrak{A}$ is exactly $(E_j\mathfrak{B})_2$. In particular, $\dim(E_j\mathfrak{B}/(E_j\mathfrak{B})_2) = \dim(\bar{e}_j\mathfrak{A})$.

For each r , $1 \leq r < p^{ns}$, the multiplication-action of M_r sends $E_j\mathfrak{B}$ into $(E_j\mathfrak{B})_{r+1}$ and induces a map of $(E_j\mathfrak{B})/(E_j\mathfrak{B})_2$ onto $(E_j\mathfrak{B})_{r+1}/(E_j\mathfrak{B})_{r+2}$. (Here $(E_j\mathfrak{B})_{p^{ns}+1} = (0)$, by definition.) Thus

$$\dim((E_j\mathfrak{B})_r/(E_j\mathfrak{B})_{r+1}) \leq \dim \bar{e}_j\mathfrak{A}$$

for every r , every j . Summing over r and j gives

$$\dim \mathfrak{B} = \sum_{r,j} \dim((E_j\mathfrak{B})_r/(E_j\mathfrak{B})_{r+1}) \leq p^{ns} \dim \mathfrak{A} = \dim \mathfrak{B},$$

yielding equality at all stages. That is, we have proved:

Proposition 1. *With notation as above, the filtration $(E_j\mathfrak{B}) \supset (E_j\mathfrak{B})_2 \supset \dots \supset (E_j\mathfrak{B})_{p^{ns}} \supset 0$ of $E_j\mathfrak{B}$ by right ideals has successive quotients which, as \mathfrak{A} -modules, are all isomorphic to $\bar{e}_j\mathfrak{A}$.*

Clearly the radical \mathfrak{N} of the algebra \mathfrak{B} contains all $\psi(T_i)$. So $E_j\mathfrak{B}/E_j\mathfrak{N}$ is a homomorphic image of $E_j\mathfrak{B}/(E_j\mathfrak{B})_2$, and the submodules of $E_j\mathfrak{B}/E_j\mathfrak{N}$ are in correspondence with the set of submodules of $\bar{e}_j\mathfrak{A}$ containing $\bar{e}_j\varphi(\mathfrak{N})$. Because $\varphi(\mathfrak{N})$ is a nilpotent ideal, $\bar{e}_j\varphi(\mathfrak{N}) \neq \bar{e}_j\mathfrak{A}$. If \bar{e}_j is a primitive idempotent, $\bar{e}_j\mathfrak{A}$ has a unique maximal submodule ([C-R], §54), *a fortiori* a unique maximal submodule containing $\bar{e}_j\varphi(\mathfrak{N})$, and, by the same reference, this submodule is $\bar{e}_j\mathfrak{A}$, where \mathfrak{A} is the radical of \mathfrak{A} . Thus $E_j\mathfrak{B}/E_j\mathfrak{N}$ has a unique maximal submodule. Because every maximal submodule of $E_j\mathfrak{B}$ must contain $E_j\mathfrak{N}$, $E_j\mathfrak{B}$ has a unique maximal submodule, and E_j is in turn a primitive idempotent. The converse is clear: If E_j is primitive, $E_j\mathfrak{B}$ has a unique maximal submodule, which must contain $(E_j\mathfrak{B})_2$; so $\bar{e}_j\mathfrak{A}$ has a unique maximal submodule. That is,

Proposition 2. *The idempotent \bar{e}_j is primitive in \mathfrak{A} if and only if E_j is primitive in \mathfrak{B} .*

To apply the above to $sl(2)$, with $z_i = h^p - h, e^p, f^p$ and the elements $\{e_j\}$ that are polynomials in the $h^*(\alpha)ef$ as in §§1-4, so that \mathfrak{B} is the quotient of $\mathfrak{U}(\mathfrak{g})$ by the ideal generated by the $(h^p - h - \lambda)^{p^s}, (e^p - \mu)^{p^s}, (f^p - \nu)^{p^s}$, we find that the images

in \mathfrak{B} of the $e_j^{p^s}$ are a system of primitive orthogonal idempotents of sum 1. In the semisimple case, each of the corresponding projective indecomposable \mathfrak{B} -modules has a composition series of length p^{3s} with quotients all isomorphic to the irreducible \mathfrak{A} -module $\bar{e}_j \mathfrak{A}$. In the non-restricted case with $\lambda^2 = 4\mu\nu$, there are p projective indecomposable modules with composition series of length p^{3s} and quotients isomorphic to the $E(\alpha, 0)\mathfrak{A}$ ($\alpha \in \mathbf{F}_p$), while the remaining p.i.m.s have composition-length $2p^{3s}$, composition-factors all isomorphic to those of $E(\alpha, \beta)\mathfrak{A}$, $\beta \in \mathbf{F}_p^{*2}$. These p.i.m.s admit a filtration where successive quotients (p^{3s} of them) are all isomorphic to $E(\alpha, \beta)\mathfrak{A}$.

In the restricted case, there are p primitive idempotents with generated modules having composition series of length p^{3s} and the Steinberg module as quotient. The remaining $\frac{p(p-1)}{2}$ primitive idempotents each yield p.i.m.s of length $4p^{3s}$, dimension $2p^{3s+1}$, with filtrations having quotients all isomorphic to the $E(\alpha, \beta)\mathfrak{A}$ as in §4. These conclusions recover Bendel's results in [Ben] (§8).

It may be noted that, when F is algebraically closed, every finite-dimensional indecomposable $sl(2)$ -module is a module for one of our algebras \mathfrak{B} , with suitable values of λ, μ, ν and s .

REFERENCES

- [Ben] C. Bendel, *Generalized reduced enveloping algebras for restricted Lie algebras*, Journal of Algebra **218** (1999), 373-411. MR **2000h**:17007
- [Ber] A. Berkson, *The u -algebra of a restricted Lie algebra is Frobenius*, Proc. Amer. Math. Soc. **15** (1964), 14-15. MR **28**:2132
- [C-R] C. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, Interscience (Wiley), New York (1962). MR **26**:2519
- [J] N. Jacobson, *Lie Algebras*, Interscience Tracts in Pure and Applied Mathematics, No. 10, Interscience (Wiley), New York (1962); Dover edition, Dover Publications, New York 1979. MR **26**:1345; MR **80k**:17001
- [JCJ] J. C. Jantzen, *Representations of Lie algebras in prime characteristic*, In *Representation Theories and Algebraic Geometry*, A. Broer and A. Daigneault, eds., Kluwer, Dordrecht/Boston/London (1998), 185-235. MR **99h**:17026
- [N] G. M. Nielsen, *A Determination of the Minimal Right Ideals in the Enveloping Algebra of a Lie Algebra of Classical Type*, Ph.D. dissertation, Madison, Wisconsin, 1963.
- [P] R. D. Pollack, *Restricted Lie algebras of bounded type*, Bull. Amer. Math. Soc. **74** (1968), 326-331. MR **36**:2661
- [S] J. Schue, *Symmetry for the enveloping algebra of a restricted Lie algebra*, Proc. Amer. Math. Soc. **16** (1965), 1123-1124. MR **32**:2515

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, P.O. BOX 208283, NEW HAVEN, CONNECTICUT 06520-8283

E-mail address: `selig@math.yale.edu`

STABILITY OF INFINITE-DIMENSIONAL SAMPLED-DATA SYSTEMS

HARTMUT LOGEMANN, RICHARD REBARBER, AND STUART TOWNLEY

ABSTRACT. Suppose that a static-state feedback stabilizes a continuous-time linear infinite-dimensional control system. We consider the following question: if we construct a sampled-data controller by applying an idealized sample-and-hold process to a continuous-time stabilizing feedback, will this sampled-data controller stabilize the system for all sufficiently small sampling times? Here the state space X and the control space U are Hilbert spaces, the system is of the form $\dot{x}(t) = Ax(t) + Bu(t)$, where A is the generator of a strongly continuous semigroup on X , and the continuous time feedback is $u(t) = Fx(t)$. The answer to the above question is known to be “yes” if X and U are finite-dimensional spaces. In the infinite-dimensional case, if F is not compact, then it is easy to find counterexamples. Therefore, we restrict attention to compact feedback. We show that the answer to the above question is “yes”, if B is a bounded operator from U into X . Moreover, if B is unbounded, we show that the answer “yes” remains correct, provided that the semigroup generated by A is analytic. We use the theory developed for static-state feedback to obtain analogous results for dynamic-output feedback control.

1. INTRODUCTION

Consider the following system with state space X and input space U (both Hilbert spaces):

$$(1.1) \quad \dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x^0 \in X,$$

where A is the generator of a strongly continuous semigroup $T(t)$ on X and B is a bounded linear operator from U into X_{-1} , where X_{-1} is a certain extrapolation space with $X \hookrightarrow X_{-1}$. If the input (or control) operator B maps boundedly into the state space X , then B is called bounded; otherwise B is called “unbounded” (with respect to the state space X). We give precise details of the set-up in Section 2.

We consider the following fundamental problem: Suppose that the feedback control $u(t) = Fx(t)$, where F is a bounded linear operator from X into U , is an exponentially stabilizing state feedback control for (1.1) in the sense that $A + BF$ generates an exponentially stable strongly continuous semigroup on X . A natural implementation of this continuous-time control $u(t) = Fx(t)$ is to use sample and

Received by the editors December 21, 2000 and, in revised form, February 21, 2002.

2000 *Mathematics Subject Classification.* Primary 34G10, 47A55, 47D06, 93C25, 93C57, 93D15.

This work was supported by NATO (Grant CRG 950179) and by the NATIONAL SCIENCE FOUNDATION (Grant DMS-9623392).

hold, so that u is given instead by

$$(1.2) \qquad u(t) = Fx(k\tau), \quad t \in [k\tau, (k+1)\tau).$$

Here $\tau > 0$ is the sampling period and $t = n\tau$, $n = 0, 1, \dots$, are the sampling times. The control (1.2) is called a sampled-data feedback control and the overall system given by (1.1) and (1.2) is called a sampled-data feedback system. Intuitively, we would expect that for all sufficiently small sampling periods $\tau > 0$, (1.2) produces a stabilizing control for (1.1) in the sense that there exist $N \geq 1$ and $\nu > 0$ so that

$$(1.3) \qquad \|x(t)\| \leq Ne^{-\nu t}\|x^0\|, \quad \forall x^0 \in X, \forall t \geq 0.$$

Integrating (1.1) and (1.2) over one sampling interval $[k\tau, (k+1)\tau)$ and setting $x_k := x(k\tau)$ yields the following discrete-time system:

$$x_{k+1} = \Delta_\tau x_k, \quad \text{where} \quad \Delta_\tau := T(\tau) + \int_0^\tau T(s)BF \, ds.$$

It is straightforward to show that Δ_τ is a bounded operator from X to X and that (1.3) holds if, and only if, Δ_τ is power stable, i.e., $\|\Delta_\tau^k\| \rightarrow 0$ as $k \rightarrow \infty$ (see Section 2). Then the fundamental problem described above becomes that of determining whether or not Δ_τ is power stable for all sufficiently small $\tau > 0$.

By way of motivation, let us briefly discuss the finite-dimensional case. Then $X = \mathbb{R}^n$, $U = \mathbb{R}^m$, A , B and F are real matrices of compatible formats and $T(t) = e^{At}$. It is instructive to sketch a simple argument showing in this case that the sampled-data feedback (1.2) is indeed stabilizing for (1.1), provided that τ is sufficiently small. Since $u(t) = Fx(t)$ is an exponentially stabilizing static state feedback control for (1.1), the eigenvalues of $A + BF$ have negative real parts, or, equivalently, there exists a positive definite symmetric matrix P satisfying

$$(1.4) \qquad (A + BF)^T P + P(A + BF) = -I.$$

Next we observe, using the power series expansion of the matrix exponential e^{At} , that there exists $M_1 \geq 0$ so that

$$\|\Delta_\tau - I - \tau(A + BF)\| \leq M_1 \tau^2, \quad \forall \tau \in (0, 1].$$

Then, invoking (1.4), it is easy to see that there exist $M_2 > 0$ and $\tau^* > 0$ such that $x^T \Delta_\tau^T P \Delta_\tau x - x^T P x \leq 2\tau x^T (A + BF) P x + M_2 \tau^2 \|x\|^2 \leq -\frac{1}{2}\tau \|x\|^2$, $\forall \tau \in (0, \tau^*)$.

Therefore, for every $\tau \in (0, \tau^*)$, $\Delta_\tau^T P \Delta_\tau - P < 0$, or, equivalently, Δ_τ is power stable, showing that the sampled-data feedback (1.2) is stabilizing for all sufficiently small $\tau > 0$. A different proof of this result can be found in Chen and Francis [1], where, using the power series expansion of the matrix exponential and the continuous dependence of the eigenvalues of Δ_τ on τ , it is shown that the spectral radius of Δ_τ is less than one for all sufficiently small $\tau > 0$ (which implies power-stability of Δ_τ for all sufficiently small $\tau > 0$).

Whilst this simple observation is easy to prove and well known in the finite-dimensional control literature, it serves to indicate that extending the above result to infinite-dimensional situations is far from trivial. Indeed, neither the proof sketched above nor the proof given in [1] have meaningful generalisations to the infinite-dimensional case, because they both rely on the convergence of the Taylor series of e^{At} in the operator-norm topology, which extends to the infinite-dimensional case only if the generator of the semigroup is bounded, or, equivalently,

if the semigroup is uniformly continuous. Systems described by uniformly continuous semigroups are not general enough to encompass most interesting examples.

It might appear that the case considered above of sampling a static state or output feedback control, whilst mathematically significant and simple to describe, is too restrictive for meaningful applications. However, results for the more practically relevant case of sampling a dynamic output feedback control can be obtained from the state feedback case quite easily, as we show in Section 5 (in an infinite-dimensional context). Such sampling of dynamic continuous-time controllers forms the basis for a so-called indirect sampled-data design methodology (see Chen and Francis [2] and Franklin, Powell and Workman [6] for a fuller description of this methodology in the finite-dimensional case).

Recently there has been increasing interest in sampled-data control of infinite-dimensional systems; see Rebarber and Townley [15], [16], Rosen and Wang [18] and Tarn et al. [20]. From these papers a trend emerges: if the underlying semigroup system is analytic, or the input operator is bounded, then typically the finite-dimensional sampled-data result generalises to the infinite-dimensional case; if the input operator is unbounded and the underlying semigroup is not analytic, then the generalisation fails. It is not surprising that analyticity plays such a key role. In some sense, analyticity constrains the spectrum of the generator of the underlying semigroup so as to avoid any losses of high-frequency information when sampling.

We give a brief description of the results contained in this paper. In Section 2 we introduce the class of infinite-dimensional systems to which our results apply and we present some preliminary basic technical results to be used in the rest of the paper. In Section 3 we consider the case in which the input operator B is bounded, but the semigroup generated by A is not assumed to be analytic. We show that if the feedback operator F is compact and the semigroup generated by $A + BF$ is exponentially stable, then, for sufficiently small $\tau > 0$, the sampled-data feedback (1.2) is stabilizing. A key element of the proof is the introduction of a new, equivalent norm on the state space with respect to which the operator Δ_τ is a strict contraction, provided that τ is sufficiently small. We also give a simple example which shows that the result is not true if the compactness assumption on F is removed. In Section 4 we consider the case in which the input operator B is unbounded and the semigroup generated by A is analytic. We show that for any compact feedback operator F , the operator $A + BF$ (with suitable domain) generates an analytic semigroup on the state space X , and if this semigroup is exponentially stable, then the sampled-data feedback (1.2) is stabilizing for all sufficiently small $\tau > 0$. The proof of this result is quite involved and differs considerably from the argument leading to the result for bounded B . Instead of trying to estimate the spectral radius of Δ_τ directly, we invoke frequency-domain stability criteria for the power stability of Δ_τ , making repeated use of the analyticity of the underlying semigroup. Finally, in Section 5 we show how to use the static state feedback results of Sections 3 and 4 in the practically more relevant situation of sampling a dynamic output feedback controller.

Notation: For $\alpha \in \mathbb{R}$, $r > 0$ and $z \in \mathbb{C}$, we define

$$\mathbb{C}_\alpha := \{s \in \mathbb{C} \mid \operatorname{Re} s > \alpha\}, \quad \mathbb{B}(z, r) := \{s \in \mathbb{C} \mid |s - z| < r\}.$$

The exterior of the unit disc is denoted by \mathbb{E} , i.e.,

$$\mathbb{E} := \{s \in \mathbb{C} \mid |s| > 1\}.$$

Let X and Y be Banach spaces. The space of all bounded linear operators from X to Y is denoted by $\mathcal{B}(X, Y)$; we set $\mathcal{B}(X) := \mathcal{B}(X, X)$. Let A be an operator with domain and range in X . Then $D(A)$ denotes the domain of A and $\varrho(A)$ denotes the resolvent set of A , i.e., the set of all $s \in \mathbb{C}$ such that $sI - A : D(A) \rightarrow X$ is bijective and $(sI - A)^{-1} \in \mathcal{B}(X)$. The spectrum of A , which is the complement of $\varrho(A)$, is denoted by $\sigma(A)$. Let $\sigma_p(A) \subset \sigma(A)$ denote the point spectrum (i.e., the set of all eigenvalues) of A . If $S \subset X$ and $A : D(A) \subset X \rightarrow Y$ is an operator, then $A|_S$ denotes the restriction of A to S , i.e., the operator defined by $A|_S x = Ax$ for all $x \in D(A|_S) := D(A) \cap S$. For an open set $\Omega \subset \mathbb{C}$, $H^\infty(\Omega, \mathcal{B}(X, Y))$ denotes the set of all bounded analytic $\mathcal{B}(X, Y)$ -valued functions f defined on an open set $\Omega_f \subset \Omega$ (depending on f) such that $\Omega \setminus \Omega_f$ is a discrete set.¹ Clearly, each function $f \in H^\infty(\Omega, \mathcal{B}(X, Y))$ admits an extension $\tilde{f} \in H^\infty(\Omega, \mathcal{B}(X, Y))$ defined on Ω . For an arbitrary set $\Lambda \subset \mathbb{C}$, we define

$$H^\infty(\Lambda, \mathcal{B}(X, Y)) := \bigcup_{\Omega \supset \Lambda, \Omega \text{ open}} H^\infty(\Omega, \mathcal{B}(X, Y)).$$

2. THE SAMPLED-DATA SYSTEM

We start with a simple lemma which will be useful in the subsequent developments.

Lemma 2.1. *Let X, Y and Z be Hilbert spaces and let $L : [0, 1] \rightarrow \mathcal{B}(X, Y)$ be given. If $\lim_{t \rightarrow 0} L(t)x = 0$ for all $x \in X$ and if $K \in \mathcal{B}(Z, X)$ is compact, then*

$$(2.1) \quad \lim_{t \rightarrow 0} \|L(t)K\|_{\mathcal{B}(Z, Y)} = 0.$$

Moreover, if $\lim_{t \rightarrow 0} L^*(t)y = 0$ for all $y \in Y$, and if $H \in \mathcal{B}(Y, Z)$ is compact, then

$$(2.2) \quad \lim_{t \rightarrow 0} \|HL(t)\|_{\mathcal{B}(X, Z)} = 0.$$

Proof. Assume that $\lim_{t \rightarrow 0} L(t)x = 0$ for all $x \in X$. If $K \in \mathcal{B}(Z, X)$ has finite rank, then it is easy to prove that (2.1) holds. In a Hilbert space a compact operator can be uniformly approximated by finite rank operators, and a standard argument shows that (2.1) is true for compact $K \in \mathcal{B}(Z, X)$.

Assuming that $\lim_{t \rightarrow 0} L^*(t)x = 0$ for all $x \in X$ and that $H \in \mathcal{B}(Y, Z)$ is compact, (2.2) can be obtained by an application of (2.1) to $L^*(t)$ and H^* . \square

Throughout the paper, X and U denote Hilbert spaces. Let us consider the following sampled-data system with state space X and input space U :

$$(2.3a) \quad \dot{x}(t) = Ax(t) + Bu(t), \quad t \geq 0; \quad x(0) = x^0 \in X,$$

$$(2.3b) \quad u(t) = Fx(k\tau), \quad k\tau \leq t < (k+1)\tau, \quad k \in \mathbb{N}_0,$$

where $x(t) \in X$, $u(t) \in U$, $\tau > 0$ is the sampling time, A is the generator of a strongly continuous semigroup $T(t)$ on X , $F \in \mathcal{B}(X, U)$ and $B \in \mathcal{B}(U, X_{-1})$, where X_{-1} denotes the closure of X in the norm $\|x\|_{-1} = \|(\lambda I - A)^{-1}x\|$ (here $\|\cdot\|$ denotes the norm of X and λ is any element in $\varrho(A)$). Clearly, $X \hookrightarrow X_{-1}$, and hence $\mathcal{B}(U, X) \subset \mathcal{B}(U, X_{-1})$. We say that B is *bounded* if $B \in \mathcal{B}(U, X)$; otherwise we say that B is *unbounded*. The semigroup $T(t)$ extends to a strongly continuous semigroup on X_{-1} . The generator of $T(t)$ on X_{-1} is an extension of A to X

¹ This means that either $\Omega_f = \Omega$ or, if $\Omega_f \neq \Omega$, then for each $z \in \Omega \setminus \Omega_f$ there exists $\varepsilon > 0$ such that $\mathbb{B}(z, \varepsilon) \cap (\Omega \setminus \Omega_f) = \{z\}$.

(which is bounded as an operator from X to X_{-1}). We shall use the same symbol $T(t)$ (respectively, A) for the original semigroup (respectively, its generator) and the associated extensions. With this convention, we may write $A \in \mathcal{B}(X, X_{-1})$. Considered as a generator on X_{-1} , the domain of A is X . The spectrum of A considered as an operator on X coincides with the spectrum of A considered as an operator on X_{-1} ; moreover, the point spectra of A and its extension coincide, including algebraic multiplicities of isolated eigenvalues. We refer the reader to p. 123 in the book [5] by Engel and Nagel for more details on the extrapolation space X_{-1} .

The derivative on the left-hand side of (2.3a) has to be interpreted in the space X_{-1} . To solve the initial-value problem (2.3), we define a function x recursively by

$$(2.4a) \quad x(0) = x^0,$$

$$(2.4b) \quad x(k\tau + t) = T(t)x(k\tau) + \int_0^t T(s)BFx(k\tau) ds, \quad \forall t \in (0, \tau], \quad \forall k \in \mathbb{N}_0.$$

It follows from standard results in the theory of strongly continuous semigroups (applied to $T(t)$ considered as a semigroup on X_{-1}) that

$$x \in C(\mathbb{R}_+, X) \quad \text{and} \quad x|_{[k\tau, (k+1)\tau]} \in C^1([k\tau, (k+1)\tau], X_{-1}), \quad \forall k \in \mathbb{N}_0.$$

Moreover, x satisfies the following differential equations in X_{-1} :

$$\dot{x}(t) = Ax(t) + BFx(k\tau), \quad \forall t \in (k\tau, (k+1)\tau), \quad \forall k \in \mathbb{N}_0.$$

It is also clear that x given by (2.4) is the only function with these properties. In this sense, the function x defined by (2.4) is the unique solution of (2.3).

We say that (2.3) is *exponentially stable* if there exist $N \geq 1$ and $\nu > 0$ such that $\|x(t)\| \leq Ne^{-\nu t}\|x^0\|$ for all initial values $x^0 \in X$ and all $t \geq 0$. To study the stability properties of the sampled-data system (2.3) we consider a related discrete time system. Let x be the solution of (2.3) given by (2.4) and set

$$x_k := x(k\tau).$$

Then, by (2.4b),

$$(2.5) \quad x_{k+1} = T(\tau)x_k + \int_0^\tau T(s)BFx_k ds = (T(\tau) + S_\tau F)x_k, \quad \forall k \in \mathbb{N}_0,$$

where S_τ is an operator defined on U and given by

$$(2.6) \quad S_\tau u := \int_0^\tau T(s)Bu ds, \quad \forall u \in U.$$

Lemma 2.2. *For any $\tau \geq 0$, $S_\tau \in \mathcal{B}(U, X)$, and for any $\theta > 0$,*

$$(2.7) \quad \sup_{0 \leq \tau \leq \theta} \|S_\tau\|_{\mathcal{B}(U, X)} < \infty.$$

Moreover, if $F \in \mathcal{B}(X, U)$ is compact, then

$$(2.8) \quad \lim_{\tau \rightarrow 0} \|S_\tau F\|_{\mathcal{B}(X)} = 0.$$

Proof. By standard results from semigroup theory applied to $T(t)$ considered as a semigroup on X_{-1} , we know that $S_\tau U \subset X$ and

$$AS_\tau u = (T(\tau) - I)Bu, \quad \forall u \in U, \quad \forall \tau \in (0, \infty).$$

Hence, for $\lambda \in \varrho(A)$, we obtain

$$(2.9) \quad \begin{aligned} S_\tau u &= (I - T(\tau))(\lambda I - A)^{-1}Bu \\ &+ \lambda \int_0^\tau T(s)(\lambda I - A)^{-1}Bu \, ds, \quad \forall u \in U, \forall \tau \in [0, \infty), \end{aligned}$$

showing that $S_\tau \in \mathcal{B}(U, X)$ and that (2.7) holds. Moreover, let $F \in \mathcal{B}(X, U)$ be compact. We see from (2.9) that $\lim_{\tau \rightarrow 0} \|S_\tau u\| = 0$ for all $u \in U$, and thus (2.8) follows from an application of Lemma 2.1. \square

Introducing the operator

$$(2.10) \quad \Delta_\tau := T(\tau) + S_\tau F,$$

(2.5) can be written in the form

$$(2.11) \quad x_{k+1} = \Delta_\tau x_k, \quad \forall k \in \mathbb{N}.$$

Note that by Lemma 2.2, $\Delta_\tau \in \mathcal{B}(X)$. Recall that an operator $L \in \mathcal{B}(X)$ is said to be *power stable* if $\|L^k\| \rightarrow 0$ as $k \rightarrow \infty$.

The following simple result can be obtained by a slight modification of the proof of Proposition 2.1 in Rebarber and Townley [16].

Lemma 2.3. *For any $\tau > 0$, (2.3) is exponentially stable if and only if Δ_τ is power stable.*

Hence, for the remainder of this paper, we study the power stability of Δ_τ in order to study the exponential stability of (2.3).

3. BOUNDED CONTROL

In this section we assume that the control operator B is bounded. We show that if the feedback operator F is compact, then exponential stability of the continuous-time semigroup generated by $A + BF$ implies exponential stability of the sampled-data system (2.3) provided the sampling time τ is small enough. More precisely, we prove the following result.

Theorem 3.1. *Assume that A generates a strongly continuous semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X)$, $F \in \mathcal{B}(X, U)$ is compact and the semigroup generated by $A + BF$ is exponentially stable. Then there exists $\tau^* > 0$ such that for every $\tau \in (0, \tau^*)$, there exist $N \geq 1$ and $\nu > 0$ such that all solutions of (2.3) satisfy $\|x(t)\| \leq Ne^{-\nu t} \|x^0\|$ for all $x^0 \in X$ and all $t \geq 0$.*

Proof. Denote the semigroup generated by $A + BF$ by $T_{BF}(t)$. We start by writing Δ_τ as a perturbation of $T_{BF}(\tau)$. By the variation of parameters formula from the perturbation theory of semigroups (see, for example, [14], p. 79), we have

$$T_{BF}(\tau)x = T(\tau)x + \int_0^\tau T(\tau - s)BFT_{BF}(s)x \, ds, \quad \forall x \in X.$$

Using (2.10) and defining $P_\tau \in \mathcal{B}(X)$ by

$$P_\tau x := \int_0^\tau T(\tau - s)BF(I - T_{BF}(s))x \, ds, \quad \forall x \in X,$$

we see that

$$(3.1) \quad \Delta_\tau = T_{BF}(\tau) + P_\tau.$$

By the hypothesis that T_{BF} is exponentially stable, there exist $M \geq 1$ and $\omega > 0$ such that

$$\|T_{BF}(t)x\| \leq Me^{-\omega t}\|x\|, \quad \forall x \in X, \quad \forall t \in \mathbb{R}_+.$$

Adopting a standard technique from semigroup theory (see, for example, the proof of Theorem 5.2 on p. 19 in Pazy [14]), we introduce a new norm $|\cdot|$ on X by setting

$$|x| := \sup_{t \geq 0} \|e^{\omega t} T_{BF}(t)x\|, \quad \forall x \in X.$$

Clearly,

$$(3.2) \quad \|x\| \leq |x| \leq M\|x\|, \quad \forall x \in X,$$

and, moreover,

$$(3.3) \quad \begin{aligned} |T_{BF}(t)x| &= \sup_{s \geq 0} \|e^{\omega s} T_{BF}(s) T_{BF}(t)x\| \leq e^{-\omega t} \sup_{s \geq 0} \|e^{\omega s} T_{BF}(s)x\| \\ &= e^{-\omega t} |x|, \quad \forall x \in X, \quad \forall t \geq 0. \end{aligned}$$

Since F is compact and $\lim_{s \rightarrow 0} (I - T_{BF}(s))x = 0$ for all $x \in X$, an application of Lemma 2.1 yields

$$(3.4) \quad \lim_{s \rightarrow 0} \|F(I - T_{BF}(s))\|_{\mathcal{B}(X, U)} = 0.$$

For $L \in \mathcal{B}(X)$ we denote the operator norm $\sup_{x \in X} (|Lx|/|x|)$ by $|L|$. Using (3.1)-(3.3) and invoking the inequality

$$e^{-\omega \tau} \leq 1 - \omega \tau e^{-\omega \tau}, \quad \forall \tau \in \mathbb{R}_+,$$

we obtain

$$(3.5) \quad \begin{aligned} |\Delta_\tau| &\leq e^{-\omega \tau} + |P_\tau| \leq e^{-\omega \tau} + M\|P_\tau\| \\ &\leq 1 - \omega \tau + [\omega(1 - e^{-\omega \tau}) + h(\tau)]\tau, \quad \forall \tau \in \mathbb{R}_+, \end{aligned}$$

where

$$h(\tau) := M \sup_{0 \leq s \leq \tau} \|T(\tau - s)BF(I - T_{BF}(s))\|_{\mathcal{B}(X)}.$$

By (3.4),

$$\lim_{\tau \rightarrow 0^+} h(\tau) = 0,$$

and hence it follows from (3.5) that for fixed but arbitrary $\varepsilon \in (0, \omega)$ there exists $\tau_\varepsilon > 0$ such that

$$(3.6) \quad |\Delta_\tau| < 1 - (\omega - \varepsilon)\tau < 1, \quad \forall \tau \in (0, \tau_\varepsilon).$$

Applying (3.2), we see that for any $\tau \in (0, \tau_\varepsilon)$,

$$\|\Delta_\tau^k x\| \leq |\Delta_\tau^k x| \leq |\Delta_\tau^k| |x| \leq M |\Delta_\tau|^k \|x\|, \quad \forall x \in X, \quad \forall k \in \mathbb{N}.$$

Combining this with (3.6) shows that Δ_τ is power stable for all $\tau \in (0, \tau_\varepsilon)$. \square

Remark 3.2. (1) Examining the proof of Theorem 3.1, we see that for every $\varepsilon \in (0, \omega)$, there exists $\tau_\varepsilon > 0$ such that

$$\|\Delta_\tau^k\| \leq M(1 - (\omega - \varepsilon)\tau)^k, \quad \forall k \in \mathbb{N}, \quad \forall \tau \in (0, \tau_\varepsilon).$$

(2) A key element in the proof of Theorem 3.1 is that $\lim_{\tau \rightarrow 0^+} h(\tau) = 0$, which implies that

$$\lim_{\tau \rightarrow 0^+} \frac{\|P_\tau\|}{\tau} = 0.$$

In the case of unbounded B , this does not seem to be true, even if B is an admissible control operator for $T(t)$ (in the sense of Weiss [21]). \diamond

The next remark shows that the compactness assumption on F imposed in Theorem 3.1 cannot be relaxed.

Remark 3.3. It is easy to find counterexamples which show that, in general, Theorem 3.1 is not true for non-compact feedback operators F . Indeed, consider the case where $X = l^2(\mathbb{Z})$, $A = \text{diag}_{k \in \mathbb{Z}}(1 + ki)$, $B = I$ and $F = -2I$. Then F is a continuous-time exponentially stabilizing feedback, since $A + BF$ generates the exponentially stable semigroup

$$T_{BF}(t) = \text{diag}_{k \in \mathbb{Z}}(e^{(-1+ki)t}).$$

However, applying the sampled-data feedback given by (2.3b) results in the discrete-time operator

$$\Delta_\tau = \text{diag}_{k \in \mathbb{Z}}(e^{(1+ki)\tau}) - 2\text{diag}_{k \in \mathbb{Z}}\left(\frac{e^{(1+ki)\tau} - 1}{1 + ki}\right).$$

Since

$$\left|e^{(1+ki)\tau} - 2\left(\frac{e^{(1+ki)\tau} - 1}{1 + ki}\right)\right| \geq e^\tau - \frac{2}{|1 + ki|}(e^\tau + 1),$$

we see that for any $\tau > 0$, the eigenvalues λ_k^τ of Δ_τ satisfy

$$\liminf_{|k| \rightarrow \infty} |\lambda_k^\tau| \geq e^\tau > 1,$$

showing that Δ_τ is not power stable for any $\tau > 0$.

Furthermore, it is possible to find counterexamples which show that, in general, Theorem 3.1 is not true when B is not bounded. In particular, in [17] an example is given for which U is one-dimensional, F is bounded, B is not bounded, $A + BF$ generates an exponentially stable semigroup, $A^{-\varepsilon}B \in \mathcal{B}(U, X)$ for any $\varepsilon > 0$ (so B is “barely unbounded” by a common measure of unboundedness), and there exists a sequence of positive sampling times (τ_n) such that $\tau_n \rightarrow 0$ and Δ_{τ_n} is not power stable. \diamond

Remark 3.4. It is important to realise that Theorem 3.1 is completely general in the sense that it applies to any generator A , any bounded control operator B and, in particular, any compact stabilizing feedback F . Obviously, in specific cases, the result could be proved using different, possibly simpler arguments. As an obvious example, consider the case of a system with a finite-dimensional unstable part (typically arising from a parabolic PDE) and a feedback that stabilizes the unstable part and is zero on the stable invariant subspace. More precisely, in this case there exists a projection operator Π on X such that $X_1 := \Pi X$ (the unstable subspace) and $X_2 := (I - \Pi)(X)$ (the stable subspace) are invariant with respect to the underlying semigroup, X_1 is finite-dimensional and $X_1 \subset D(A)$; moreover,

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix},$$

where $A_1 := A|_{X_1}$, $A_2 := A|_{X_2}$, $B_1 := \Pi B$, $B_2 := (I - \Pi)B$ and the operator A_2 generates an exponentially stable semigroup on X_2 (see, for example, Lemma 2.5.7 in Curtain and Zwart [3] for more details). If $F_1 : X_1 \rightarrow U$ is an exponentially stabilizing feedback for the (finite-dimensional) system (A_1, B_1) , then $F := (F_1, 0)$

is an exponentially stabilizing feedback for (A, B) . It is obvious from the well-known finite-dimensional version of Theorem 3.1 that for this particular feedback the corresponding sampled-data feedback system (2.3) is exponentially stable for all sufficiently small $\tau > 0$. Clearly, one could imagine proofs tailored to other situations with specific feedbacks, say for example hyperbolic PDEs, retarded and neutral functional equations or mixed parabolic/hyperbolic systems. However, in many significant applications there is little scope for actually choosing the feedback. For example, in static output feedback, the basic structure of the feedback is inherited from the given observation and the feedback design is limited to choosing feedback gains. In such cases it is difficult to see the merits of specifically tailored proofs or indeed how they would be developed. It is therefore an important aspect of Theorem 3.1 that it does not make any assumptions on the specific structure of the underlying system (A, B) and the stabilizing feedback F . \diamond

In the next section we show that Theorem 3.1 generalises to the case with unbounded B in the special, but important, case where A generates an analytic semigroup.

4. UNBOUNDED CONTROL AND ANALYTIC SEMIGROUP

In order to formulate and prove the main result of this section a number of preparations are required. Throughout this section, we assume that A generates a strongly continuous semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X_{-1})$ and $F \in \mathcal{B}(X, U)$. Further assumptions on (A, B, F) will be introduced when needed. We define an operator A_{BF} on X by setting

$$(4.1) \quad A_{BF}x = (A + BF)x, \quad \forall x \in D(A_{BF}) := \{x \in X \mid (A + BF)x \in X\}.$$

It is easy to show that A_{BF} is closed. We carefully distinguish between the operators A_{BF} and $A + BF$, the latter being an unbounded operator on X_{-1} with domain $D(A + BF) = X$. Clearly, $A + BF \in \mathcal{B}(X, X_{-1})$. Note that A_{BF} is the part of $A + BF$ in X . The transfer function of the continuous-time system given by (A, B, F) is defined as follows:

$$(4.2) \quad \mathbf{G} : \varrho(A) \rightarrow \mathcal{B}(U), \quad s \mapsto F(sI - A)^{-1}B.$$

We introduce the space $X_* := D(A^*)$, endowed with the norm

$$\|x\|_* := \|(\bar{\lambda}I - A^*)x\|,$$

where λ is any element of $\varrho(A)$. Clearly, $X_1 \hookrightarrow X = X^*$. It is well known that X_{-1} and $(X_*)^*$ are isometrically isomorphic. Therefore, the triple $X_* \hookrightarrow X = X^* \hookrightarrow X_{-1}$ is a so-called Gelfand triple. Moreover, if $B \in \mathcal{B}(U, X_{-1})$, then $B^* \in \mathcal{B}(X_*, U)$.

Lemma 4.1. *Under the above assumptions on A , B and F , the following statements hold:*

- (1) $\varrho(A + BF) \cap \varrho(A) = \{s \in \varrho(A) \mid 1 \in \varrho(\mathbf{G}(s))\}$;
- (2) if $\varrho(A + BF) \neq \emptyset$, then $\varrho(A_{BF}) = \varrho(A + BF)$ and

$$(sI - A_{BF})^{-1} = (sI - A - BF)^{-1}|_X, \quad \forall s \in \varrho(A_{BF}) = \varrho(A + BF);$$

- (3) if F is compact, then

$$\varrho(A_{BF}) \cap \varrho(A) \subset \varrho(A + BF);$$

- (4) if F is compact, then $A_{BF}^* = A^* + F^*B^*$ with $D(A_{BF}^*) = D(A^*) = X_*$.

Proof. To prove statements (1) and (3), let $s \in \varrho(A)$, and write

$$(4.3) \quad sI - A - BF = (sI - A)[I - (sI - A)^{-1}BF].$$

It follows that

$$s \in \varrho(A + BF) \Leftrightarrow 1 \in \varrho((sI - A)^{-1}BF) \Leftrightarrow 1 \in \varrho(F(sI - A)^{-1}B),$$

yielding statement (1). Assume that F is compact. Then, by (4.3),

$$s \in \sigma(A + BF) \Leftrightarrow 1 \in \sigma_p((sI - A)^{-1}BF).$$

Therefore, if $s \in \sigma(A + BF)$, there exists $x \in X$, $x \neq 0$, such that $(sI - A)^{-1}BFx = x$. Hence, $(A + BF)x = sx$, showing that $s \in \sigma_p(A_{BF})$. We may conclude that $\sigma(A + BF) \cap \varrho(A) \subset \sigma(A_{BF})$, proving statement (3).

To prove statement (2), let $\lambda \in \varrho(A + BF)$ and note that, by Proposition 2.17 on p. 261 in [5], we only need to show that the topology on X given by the norm

$$x \mapsto |x| = \|(\lambda I - A - BF)x\|_{-1},$$

is stronger than the original norm topology. But this follows immediately, since with

$$k := \|(\lambda I - A - BF)^{-1}\|_{\mathcal{B}(X_{-1}, X)}, \quad l := \|\lambda I - A - BF\|_{\mathcal{B}(X, X_{-1})}$$

we have

$$\|x\| \leq k\|(\lambda I - A - BF)x\|_{-1} = k|x|, \quad |x| \leq l\|x\|; \quad \forall x \in X,$$

i.e., the two norms are even equivalent.

To prove statement (4), consider the operator $A^* + F^*B^*$ with $D(A^* + F^*B^*) = D(A^*) = X_*$. The compactness of F together with the fact that $B^* \in \mathcal{B}(X_*, U)$ yields that F^*B^* is relatively compact with respect to A^* . Therefore, since A^* is closed, $A^* + F^*B^*$ is closed (see Theorem 1.1 on p. 194 in [8]) and so

$$(A^* + F^*B^*)^{**} = A^* + F^*B^*$$

(see Theorem 5.29 on p. 168 in [8]). As a consequence, statement (4) will follow if we can show that

$$(A^* + F^*B^*)^* = A_{BF}.$$

To this end assume that $x \in D((A^* + F^*B^*)^*)$. Setting $x' := (A^* + F^*B^*)^*x \in X$, we have that

$$(4.4) \quad \langle x, (A^* + F^*B^*)z \rangle = \langle x', z \rangle, \quad \forall z \in X_*,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in X . The duality pairing on $X_{-1} \times X_*$, denoted by $[\cdot, \cdot]$, is given by

$$[y, z] := \langle (\lambda I - A)^{-1}y, (\bar{\lambda}I - A^*)z \rangle, \quad \forall (y, z) \in X_{-1} \times X_*,$$

where $\lambda \in \varrho(A)$. It is straightforward to show that

$$(4.5) \quad [y, z] = \langle y, z \rangle, \quad [Ay, z] = \langle y, A^*z \rangle, \quad [BFy, z] = \langle y, F^*B^*z \rangle; \quad \forall (y, z) \in X \times X_*.$$

Thus, by (4.4),

$$[x' - (A + BF)x, z] = 0, \quad \forall z \in X_*,$$

and so $(A + BF)x = x' \in X$, proving that $x \in D(A_{BF})$ and $A_{BF}x = (A^* + F^*B^*)^*x$. This shows that A_{BF} is an extension of $(A^* + F^*B^*)^*$. It remains to prove that

$D(A_{BF}) \subset D((A^* + F^*B^*)^*)$. Let $x \in D(A_{BF})$. Invoking (4.5), we obtain for all $z \in X_*$,

$$\langle x, (A^* + F^*B^*)z \rangle = [(A + BF)x, z] = \langle A_{BF}x, z \rangle,$$

showing that $x \in D((A^* + F^*B^*)^*)$. \square

Remark 4.2. Statements (2) and (3) of Lemma 4.1 are wrong if the operator A_{BF} is replaced by a proper restriction of A_{BF} . More precisely, consider an operator \tilde{A}_{BF} on X given by

$$\tilde{A}_{BF}x = (A + BF)x, \quad \forall x \in D(\tilde{A}_{BF}) \subset D(A_{BF}).$$

If $D(\tilde{A}_{BF}) \neq D(A_{BF})$, then

$$(4.6) \quad \varrho(\tilde{A}_{BF}) \subset \sigma_p(A_{BF}) \subset \sigma_p(A + BF).$$

The second inclusion holds trivially. To prove the first inclusion, let $\lambda \in \varrho(\tilde{A}_{BF})$. Then, $\lambda I - \tilde{A}_{BF} : D(\tilde{A}_{BF}) \rightarrow X$ is bijective and hence, for $y \in D(A_{BF}) \setminus D(\tilde{A}_{BF})$, there exists $x \in D(\tilde{A}_{BF})$ such that $(\lambda I - \tilde{A}_{BF})x = (\lambda I - A_{BF})y$. It follows that $(\lambda I - A_{BF})(y - x) = 0$, and since $y - x \neq 0$, we see that $\lambda \in \sigma_p(A_{BF})$.

Moreover, combining statement (1) of Lemma 4.1 and (4.6), we see that if \tilde{A}_{BF} is a proper restriction of A_{BF} , then $I - \mathbf{G}(s)$ is not invertible for any $s \in \varrho(\tilde{A}_{BF}) \cap \varrho(A)$. \diamond

Lemma 4.3. Assume that $F \in \mathcal{B}(X, U)$ is compact and that the following two spectral assumptions hold:

(S1) $\bar{\mathbb{C}}_0 \subset \varrho(A_{BF})$;

(S2) there exists $\varepsilon > 0$ such that $\sigma(A) \cap \mathbb{C}_{-\varepsilon}$ is bounded.

Then there exists $\alpha < 0$ such that $\sigma(A) \cap \mathbb{C}_\alpha$ consists of finitely many eigenvalues of A with finite algebraic multiplicities.

Proof. Since F is compact, it follows from statement (3) of Lemma 4.1 and assumption (S1) that $\varrho(A + BF) \neq \emptyset$. Hence, combining statement (2) of Lemma 4.1, assumption (S1) and the fact that $\varrho(A_{BF})$ is open shows that there exists an open set Ω such that

$$(4.7) \quad \varrho(A + BF) = \varrho(A_{BF}) \supset \Omega \supset \bar{\mathbb{C}}_0.$$

Recall that the spectrum of A considered as an operator on X coincides with the spectrum of A considered as an operator on X_{-1} (see [5], p. 261, Proposition 2.17). Moreover, the point spectra of A and its extension are identical, including algebraic multiplicities of isolated eigenvalues. For the rest of this proof we will consider A as an operator on X_{-1} with domain X . Clearly,

$$(4.8) \quad sI - A = [I + BF(sI - A - BF)^{-1}](sI - A - BF), \quad \forall s \in \Omega,$$

and therefore, by the compactness of F ,

$$(4.9) \quad s \in \sigma(A) \Leftrightarrow -1 \in \sigma_p(BF(sI - A - BF)^{-1}), \quad \forall s \in \Omega.$$

Let $s \in \sigma(A) \cap \Omega$. Then, by (4.8) and (4.9), there exists $x \in X_{-1}$, $x \neq 0$, such that

$$(sI - A)(sI - A - BF)^{-1}x = 0,$$

showing that s is an eigenvalue of A , and so

$$(4.10) \quad \sigma(A) \cap \Omega = \sigma_p(A) \cap \Omega.$$

Combining (4.9), (4.10) and the compactness of F with Theorem 1.9 on p. 370 in Kato [8] shows that for any compact set $K \subset \Omega$, the intersection $\sigma_p(A) \cap K$ is finite. It now follows from assumption (S2) that there exists $\alpha < 0$ such that $\sigma_p(A) \cap \mathbb{C}_\alpha$ is finite and $\sigma_p(A) \cap \mathbb{C}_\alpha \subset \Omega$. It remains to show that each element in $\sigma_p(A) \cap \mathbb{C}_\alpha$ has finite algebraic multiplicity. Seeking a contradiction, let $\lambda \in \sigma_p(A) \cap \mathbb{C}_\alpha \subset \Omega$ and suppose that λ has infinite algebraic multiplicity. Then, by Theorem 5.28 on p. 239 in [8], λ belongs to the essential spectrum of A (as defined on p. 243 in [8]). As an unbounded perturbation on X_{-1} , the operator BF is relatively compact with respect to A (considered on X_{-1}), due to the compactness of F . By Theorem 5.35 on p. 244 in [8], the essential spectrum of an operator remains fixed under relatively compact perturbations, and thus, $\lambda \in \sigma(A + BF) \cap \Omega$, contradicting (4.7). \square

Assume that $F \in \mathcal{B}(X, U)$ is compact and the assumptions (S1) and (S2) hold. Then, by Lemma 4.3, there exists $\alpha < 0$ such that $\sigma(A) \cap \mathbb{C}_\alpha$ consists of finitely many eigenvalues of A with finite algebraic multiplicities. Hence, there exists a simple closed curve γ in the half-plane \mathbb{C}_α not intersecting $\sigma(A)$ and enclosing $\sigma(A) \cap \bar{\mathbb{C}}_0$ in its interior. The operator

$$(4.11) \quad \Pi := \frac{1}{2\pi i} \int_{\gamma} (sI - A)^{-1} ds$$

is a projection operator, and we have

$$(4.12) \quad X = X_1 \oplus X_2, \quad \text{where } X_1 := \Pi X, \quad X_2 := (I - \Pi)X.$$

By a standard result (see, for example, Lemma 2.5.7 in Curtain and Zwart [3]), the above decomposition has the following properties:

- (D1) $\dim X_1 < \infty$ and $X_1 \subset D(A)$;
- (D2) X_1 and X_2 are $T(t)$ -invariant for all $t \geq 0$;
- (D3) $\sigma(A|_{X_1}) = \sigma(A) \cap \bar{\mathbb{C}}_0$ and $\sigma(A|_{X_2}) = \sigma(A) \cap (\mathbb{C} \setminus \bar{\mathbb{C}}_0)$.

It is useful to introduce the following notation:

$$(4.13) \quad A_j := A|_{X_j}, \quad T_j(t) := T(t)|_{X_j} \quad j = 1, 2.$$

Clearly, by (D1) and (D2), $T_1(t)$ is a semigroup on the finite-dimensional space X_1 with generator A_1 , i.e., $T_1(t) = e^{A_1 t}$, and $T_2(t)$ is a strongly continuous semigroup on X_2 with generator A_2 . Since the spectrum of A considered as an operator on X coincides with the spectrum of A considered as an operator on X_{-1} , the projection operator Π on X defined in (4.11) extends to a projection on X_{-1} . We will use the same symbol Π for the original projection and its associated extension. The decomposition (4.12) induces decompositions of the control operator $B \in \mathcal{B}(U, X_{-1})$ and the feedback operator $F \in \mathcal{B}(X, U)$:

$$(4.14) \quad B_1 := \Pi B, \quad B_2 := (I - \Pi)B, \quad F_1 := F|_{X_1}, \quad F_2 := F|_{X_2}.$$

Lemma 4.4. *Assume that $F \in \mathcal{B}(X, U)$ is compact and that assumptions (S1) and (S2) hold. Then the pair (A_1, B_1) is controllable and the pair (A_1, F_1) is observable.*

Proof. Seeking a contradiction, suppose that (A_1, B_1) is not controllable. Then, by the Kalman controllability decomposition lemma, we may assume, without loss of generality, that A_1 and B_1 can be written in the form

$$(4.15) \quad A_1 = \begin{pmatrix} A_{11} & * \\ 0 & A_{12} \end{pmatrix}, \quad B_1 = \begin{pmatrix} B_{11} \\ 0 \end{pmatrix}.$$

A straightforward calculation combined with an application of statement (2) of Lemma 4.1 shows that

$$(4.16) \quad (sI - A_{BF})^{-1} = (sI - A)^{-1} + (sI - A)^{-1}BF(sI - A_{BF})^{-1}, \quad \forall s \in \varrho(A) \cap \bar{\mathbb{C}}_0.$$

Using the decomposition of A and B and (4.15), we see that $(sI - A)^{-1}BF$ may be written in the form

$$(4.17) \quad (sI - A)^{-1}BF = \begin{pmatrix} (sI - A_{11})^{-1} & * & 0 \\ 0 & (sI - A_{12})^{-1} & 0 \\ 0 & 0 & (sI - A_2)^{-1} \end{pmatrix} \begin{pmatrix} * & * & * \\ 0 & 0 & 0 \\ * & * & * \end{pmatrix}.$$

Let $\lambda \in \sigma(A_{12})$ and let y be a corresponding eigenvector of A_{12} . Note that, by (D3), $\lambda \in \bar{\mathbb{C}}_0$. Setting $x := \text{col}(0, y, 0) \in X$, we may conclude from (4.16) and (4.17) that

$$(sI - A_{BF})^{-1}x = \begin{pmatrix} * \\ (sI - A_{12})^{-1}y \\ 0 \end{pmatrix} + \begin{pmatrix} * \\ 0 \\ * \end{pmatrix} = \begin{pmatrix} * \\ (s - \lambda)^{-1}y \\ * \end{pmatrix}.$$

This shows that $\|(sI - A_{BF})^{-1}x\| \rightarrow \infty$ as $s \rightarrow \lambda$ in $\varrho(A) \cap \bar{\mathbb{C}}_0$, leading to a contradiction, since by (S1), $\lambda \in \varrho(A_{BF})$.

To show the observability of (A_1, F_1) , note that

$$(4.18) \quad (sI - A_{BF})^{-1} = (sI - A)^{-1} + (sI - A - BF)^{-1}BF(sI - A)^{-1}, \quad \forall s \in \varrho(A) \cap \bar{\mathbb{C}}_0,$$

which can be obtained from a straightforward calculation combined with an application of statement (2) of Lemma 4.1. Based on the Kalman observability decomposition lemma and (4.18), an argument similar to that establishing controllability of (A_1, B_1) can be used to prove observability of (A_1, F_1) . We omit the details for the sake of brevity. \square

The next result shows that if A generates an analytic semigroup and if F is compact, then A_{BF} generates an analytic semigroup.

Proposition 4.5. *Assume that A generates an analytic semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X_{-1})$, and $F \in \mathcal{B}(X, U)$ is compact. Then A_{BF} generates an analytic semigroup $T_{BF}(t)$ on X . Moreover, if $T_{BF}(t)$ is exponentially stable, then $(I - \mathbf{G})^{-1} \in H^\infty(\mathbb{C}_0, \mathcal{B}(U))$, where \mathbf{G} is given by (4.2).*

Proof. By statement (4) of Lemma 4.1, $A_{BF}^* = A^* + F^*B^*$ with $D(A_{BF}^*) = D(A^*) = X_1$. Clearly, since A generates an analytic semigroup, so does A^* . Moreover, by the assumptions on B and F , $B^* \in \mathcal{B}(X_1, U)$ and F^* is compact, and so F^*B^* is relatively compact with respect to A^* . By a well-known result from the perturbation theory of analytic semigroups (see Corollary 2.17 on p. 180 in [5] or Proposition 1 in Zabczyk [23]), it follows that $A_{BF}^* = A^* + F^*B^*$ generates an analytic semigroup, and hence, so does A_{BF}^{**} . By the closedness of A_{BF} , we have that $A_{BF}^{**} = A_{BF}$, and thus we may conclude that A_{BF} generates an analytic semigroup $T_{BF}(t)$.

Assume that $T_{BF}(t)$ is exponentially stable. Then, by Lemma 4.1,

$$\varrho(A_{BF}) = \varrho(A + BF) \supset \bar{\mathbb{C}}_0,$$

and for all $s \in \bar{\mathbb{C}}_0$ and for all $x \in X_{-1}$ we have that

$$(sI - A - BF)^{-1}x = (A + BF)(sI - A - BF)^{-1}(A + BF)^{-1}x.$$

Since $(A + BF)^{-1}x \in X$, we may conclude, using Lemma 4.1, that for all $s \in \bar{\mathbb{C}}_0$ and all $x \in X_{-1}$,

$$\begin{aligned} (sI - A - BF)^{-1}x &= A_{BF}(sI - A_{BF})^{-1}(A + BF)^{-1}x \\ (4.19) \qquad \qquad &= (s(sI - A_{BF})^{-1} - I)(A + BF)^{-1}x. \end{aligned}$$

A straightforward calculation shows that

$$(I - \mathbf{G}(s))^{-1} = (I - \mathbf{G}(s))^{-1}\mathbf{G}(s) + I = F(sI - A - BF)^{-1}B + I,$$

and so, by (4.19),

$$(4.20) \qquad (I - \mathbf{G}(s))^{-1} = F(s(sI - A_{BF})^{-1} - I)(A + BF)^{-1}B + I.$$

Since $T_{BF}(t)$ is an exponentially stable analytic semigroup and since $(A + BF)^{-1}B \in \mathcal{B}(U, X)$, the right-hand side of (4.20) is uniformly bounded for $s \in \mathbb{C}_0$. Therefore, $(I - \mathbf{G})^{-1} \in H^\infty(\mathbb{C}_0, \mathcal{B}(U))$. \square

Corollary 4.6. *Assume that A generates an analytic semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X_{-1})$, and $F \in \mathcal{B}(X, U)$ is compact. If the semigroup generated by A_{BF} is exponentially stable, then there exists a decomposition $X = X_1 \oplus X_2$ such that (D1)–(D3) are satisfied and, moreover, the following property holds:*

(D4) $T_2(t)$ is an exponentially stable analytic semigroup on X_2 with generator A_2 , where $T_2(t)$ and A_2 are given by (4.13).

Proof. By the exponential stability of the semigroup generated by A_{BF} , assumption (S1) in Lemma 4.3 holds. Since A generates an analytic semigroup, for any $\beta \in \mathbb{R}$, the intersection $\sigma(A) \cap \mathbb{C}_\beta$ is bounded, and so assumption (S2) in Lemma 4.3 is satisfied. Therefore, by Lemma 4.3, properties (D1)–(D3) hold. Moreover, it is clear that the semigroup $T_2(t)$ is analytic, and therefore satisfies the spectrum determined growth assumption. By (D3) and Lemma 4.3, the generator A_2 of $T_2(t)$ satisfies $\sigma(A_2) \subset \mathbb{C} \setminus \mathbb{C}_{-\varepsilon}$ for some $\varepsilon > 0$. Therefore, we may conclude that $T_2(t)$ is exponentially stable. \square

Consider the discrete-time system $(T(\tau), S_\tau, F)$, where S_τ is given by (2.6), and denote the transfer function of this system by \mathbf{H}_τ , i.e.,

$$(4.21) \qquad \mathbf{H}_\tau(z) = F(zI - T(\tau))^{-1}S_\tau.$$

Then, under the assumptions of Lemma 4.3 (i.e., F is compact and (S1) and (S2) hold), we may write

$$(4.22) \qquad \mathbf{H}_\tau = \mathbf{H}_\tau^1 + \mathbf{H}_\tau^2,$$

where, using the notation of (4.13) and (4.14),

$$(4.23) \qquad \mathbf{H}_\tau^j(z) := F_j(zI - T_j(\tau))^{-1} \int_0^\tau T_j(s)B_j ds, \quad j = 1, 2.$$

In light of Lemma 2.3, in order to prove that the sampled-data system (2.3) is exponentially stable for a given sampling time $\tau > 0$, we need to show that Δ_τ given by (2.10) is power stable. The following lemma provides a sufficient condition for the power stability of Δ_τ in terms of τ and \mathbf{H}_τ .

Lemma 4.7. *Assume that A generates an analytic semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X_{-1})$, $F \in \mathcal{B}(X, U)$ is compact and the semigroup generated by A_{BF} is exponentially stable. If $\tau > 0$ is such that*

$$(4.24) \qquad \tau(\lambda_2 - \lambda_1) \neq 2k\pi i, \quad \forall k \in \mathbb{Z} \setminus \{0\}, \quad \forall \lambda_1, \lambda_2 \in \sigma(A) \cap \bar{\mathbb{C}}_0,$$

and if $(I - \mathbf{H}_\tau)^{-1} \in H^\infty(\mathbb{E}, \mathcal{B}(U))$, then Δ_τ is power stable.

Note that under the assumptions of the above lemma, (4.24) is satisfied for all sufficiently small $\tau > 0$.

Proof of Lemma 4.7. By Corollary 4.6, there exists a decomposition $X = X_1 \oplus X_2$ such that (D1)–(D4) hold. In particular, using the notation of (4.13) and (4.14), we have $\sigma(A_1) = \sigma(A) \cap \bar{\mathbb{C}}_0$, and, moreover, setting $S_\tau^j := \int_0^\tau T_j(s)B_j ds$, $j = 1, 2$, we may write

$$(4.25) \quad T(\tau) = \begin{pmatrix} T_1(\tau) & 0 \\ 0 & T_2(\tau) \end{pmatrix}, \quad S_\tau = \begin{pmatrix} S_\tau^1 \\ S_\tau^2 \end{pmatrix}, \quad F = (F_1, F_2).$$

Assume that τ satisfies (4.24). By Lemma 4.4, the finite-dimensional controlled system (A_1, B_1) is controllable, and so, by a well-known result (see [7] or Theorem 4 on p. 102 in Sontag [19]), the discrete-time system $(T_1(\tau), S_\tau^1)$ is controllable as well. This implies, in particular, that there exists $K_1 \in \mathcal{B}(X_1, U)$ such that the matrix $T_1(\tau) + S_\tau^1 K_1$ is power stable. Now $T_2(\tau)$ is power stable (since the semigroup $T_2(t)$ is exponentially stable by (D4)) and, setting $K := (K_1, 0) \in \mathcal{B}(X, U)$, it follows from (4.25) that $T(\tau) + S_\tau K$ is power stable, showing that the controlled discrete-time system $(T(\tau), S_\tau)$ is stabilizable. A similar argument shows that the observed discrete-time system $(T(\tau), F)$ is detectable. Hence, if $(I - \mathbf{H}_\tau)^{-1} \in H^\infty(\mathbb{E}, \mathcal{B}(U))$, we may conclude that Δ_τ is power stable (see Theorem 2 in Logemann [9]). \square

The following theorem is the main result of this section.

Theorem 4.8. *Assume that A generates an analytic semigroup $T(t)$ on X , $B \in \mathcal{B}(U, X_{-1})$, and $F \in \mathcal{B}(X, U)$ is compact. If the semigroup generated by A_{BF} is exponentially stable, then there exists $\tau^* > 0$ such that for every $\tau \in (0, \tau^*)$, there exist $N \geq 1$ and $\nu > 0$ such that all solutions of (2.3) satisfy $\|x(t)\| \leq Ne^{-\nu t}\|x^0\|$ for all $x^0 \in X$ and all $t \geq 0$.*

We see that the above result is of the same form as Theorem 3.1. In particular, no assumptions are made on the specific structure of the stabilizing continuous-time feedback F . It is therefore clear that Remark 3.4 also applies to Theorem 4.8.

The following lemma is the key tool for the proof of Theorem 4.8.

Lemma 4.9. *For $\eta > 0$, define $W(\eta) := \{z = e^s \mid s \in \mathbb{C}_0, |s| < \eta\}$. Under the assumptions of Theorem 4.8, there exist $\eta > 0$ and $\theta > 0$ such that*

$$(4.26) \quad (I - \mathbf{H}_\tau)^{-1} \in H^\infty(W(\eta), \mathcal{B}(U)) \quad \forall \tau \in (0, \theta),$$

where \mathbf{H}_τ is given by (4.21).

We defer the long and technical proof of Lemma 4.9 to the end of this section.

Proof of Theorem 4.8. By Lemma 2.3 and Lemma 4.7, it is sufficient to show that there exists $\tau^* > 0$ such that

$$(4.27) \quad (I - \mathbf{H}_\tau)^{-1} \in H^\infty(\mathbb{E}, \mathcal{B}(U)), \quad \forall \tau \in (0, \tau^*).$$

To this end, let $M \geq 1$ and $\mu \in \mathbb{R}$ be such that

$$(4.28) \quad \|T(t)\| \leq Me^{\mu t}, \quad \forall t \in \mathbb{R}_+.$$

We split \mathbb{E} into three disjoint parts:

$$(4.29) \quad \mathbb{E} = E_1(\delta) \cup E_2(\delta) \cup E_3,$$

where, for $\delta \in (0, 1)$,

$$\begin{aligned} E_1(\delta) &:= \{z \in \mathbb{C} \mid 1 < |z| < 4M, |z - 1| \leq \delta\}, \\ E_2(\delta) &:= \{z \in \mathbb{C} \mid 1 < |z| < 4M, |z - 1| > \delta\}, \end{aligned}$$

and $E_3 := \{z \in \mathbb{C} \mid |z| \geq 4M\}$. We proceed in several steps.

Step 1: We claim that there exist $\delta > 0$ and $\tau_1 > 0$ such that

$$(4.30) \qquad (I - \mathbf{H}_\tau)^{-1} \in H^\infty(E_1(\delta), \mathcal{B}(U)), \quad \forall \tau \in (0, \tau_1).$$

By Lemma 4.9 there exist $\eta > 0$ and $\theta > 0$ such that (4.26) holds. Choosing $\delta \in (0, 1)$ such that $E_1(\delta) \subset W(\eta)$, we see that (4.30) follows with $\tau_1 = \theta$.

Step 2: Let $\delta \in (0, 1)$ be as in Step 1. We claim that there exists $\tau_2 > 0$ such that

$$(4.31) \qquad (I - \mathbf{H}_\tau)^{-1} \in H^\infty(E_2(\delta), \mathcal{B}(U)), \quad \forall \tau \in (0, \tau_2).$$

To this end, first note that by (4.23)

$$\mathbf{H}_\tau^1(z) = F_1((z - 1)I - (T_1(\tau) - I))^{-1} \int_0^\tau T_1(s)B_1 \, ds.$$

Since $T_1(t)$ is a semigroup on the finite-dimensional space X_1 , we have

$$\lim_{\tau \rightarrow 0} \|T_1(\tau) - I\| = 0.$$

Moreover,

$$\lim_{\tau \rightarrow 0} \left\| \int_0^\tau T_1(s)B_1 \, ds \right\| = 0.$$

Since $|z - 1| > \delta > 0$ for all $z \in E_2(\delta)$, a routine argument involving the application of the Neumann series shows that there there exists $\theta^* > 0$ such that

$$(4.32) \qquad \|\mathbf{H}_\tau^1(z)\| \leq 1/4, \quad \forall z \in E_2(\delta), \forall \tau \in (0, \theta^*).$$

Now we analyze $\mathbf{H}_\tau^2(z)$ on $E_2^{cl}(\delta)$, the closure of $E_2(\delta)$. Using analyticity and boundedness of the semigroup $T_2(t)$ (guaranteed by (D4)), an application of Theorem 5.6 (b) on p. 66 in Pazy [14] shows that for each $\lambda \in E_2^{cl}(\delta)$ there exist $\theta_\lambda > 0$ and $k_\lambda > 0$ such that

$$(4.33) \qquad \|(\lambda I - T_2(\tau))^{-1}\| \leq k_\lambda, \quad \forall \tau \in (0, \theta_\lambda).$$

For $z, \lambda \in E_2^{cl}(\delta)$ we have

$$\|(zI - T_2(\tau))^{-1}\| \leq \|(T_2(\tau) - \lambda I)^{-1}\| \|(I - (z - \lambda)(T_2(\tau) - \lambda I)^{-1})^{-1}\|.$$

Combining this with (4.33), an application of the Neumann series yields that for each $\lambda \in E_2^{cl}(\delta)$,

$$\|(zI - T_2(\tau))^{-1}\| \leq 2k_\lambda, \quad \forall z \in \mathbb{B}(\lambda, 1/(2k_\lambda)), \forall \tau \in (0, \theta_\lambda).$$

Since $E_2^{cl}(\delta)$ is compact, we may conclude from a standard compactness argument that there exist $k > 0$ and $\tilde{\theta} > 0$ such that

$$(4.34) \qquad \|(zI - T_2(\tau))^{-1}\| \leq k, \quad \forall z \in E_2^{cl}(\delta), \tau \in (0, \tilde{\theta}).$$

By (4.23),

$$\mathbf{H}_\tau^2(z) = F_2(T_2(\tau) - I)(zI - T_2(\tau))^{-1}A_2^{-1}B_2,$$

and since F_2 is compact, it follows from Lemma 2.1 and (4.34) that there exists $\hat{\theta} \in (0, \tilde{\theta})$ such that

$$\|\mathbf{H}_\tau^2(z)\| \leq 1/4, \quad \forall z \in E_2(\delta), \tau \in (0, \hat{\theta}).$$

Combining this with (4.32) and (4.22) shows that

$$\|\mathbf{H}_\tau(z)\| \leq 1/2, \quad \forall z \in E_2(\delta), \tau \in (0, \tau_2),$$

where $\tau_2 := \min(\theta^*, \hat{\theta})$. Yet another application of the Neumann series yields (4.31).

Step 3: We claim that there exists $\tau_3 > 0$ such that

$$(4.35) \quad (I - \mathbf{H}_\tau)^{-1} \in H^\infty(E_3, \mathcal{B}(U)), \quad \forall \tau \in (0, \tau_3).$$

By (2.10) and (4.28), together with Lemma 2.2, there exists $\tau_3 > 0$ such that

$$\|\Delta_\tau\| \leq 2M, \quad \forall \tau \in (0, \tau_3).$$

Defining the open set

$$\tilde{E}_3 := \{z \in \mathbb{C} \mid |z| > 3M\} \supset E_3,$$

a routine argument involving the Neumann series then shows that

$$(4.36) \quad \|(zI - \Delta_\tau)^{-1}\| = |z^{-1}| \|(I - z^{-1}\Delta_\tau)^{-1}\| \leq 1/M, \quad \forall z \in \tilde{E}_3, \forall \tau \in (0, \tau_3).$$

Now

$$(4.37) \quad (I - \mathbf{H}_\tau(z))^{-1} = I + F(zI - \Delta_\tau)^{-1}S_\tau,$$

and so (4.35) follows from (4.36), (4.37) and Lemma 2.2.

Step 4: Finally, setting $\tau^* = \min(\tau_1, \tau_2, \tau_3)$, a combination of (4.29)–(4.31) and (4.35) yields (4.27). \square

It remains to prove the crucial Lemma 4.9.

Proof of Lemma 4.9. Recall the definition of \mathbf{G} in (4.2). Appealing to (D1)–(D4), (4.13) and (4.14), we may write

$$(4.38) \quad \mathbf{G} = \mathbf{G}_1 + \mathbf{G}_2, \quad \text{where } \mathbf{G}_j(s) = F_j(sI - A_j)^{-1}B_j, \quad j = 1, 2.$$

It is convenient to set

$$\mathbb{C}_0^* := \mathbb{C}_0 \setminus \sigma(A) = \mathbb{C}_0 \setminus \sigma(A_1).$$

For $\eta > 0$ and $\theta > 0$, we define

$$\begin{aligned} V(\theta, \eta) &:= \{(\tau, s) \in (0, \infty) \times \mathbb{C}_0 \mid \tau \in (0, \theta), \tau|s| < \eta\}, \\ V^*(\theta, \eta) &:= \{(\tau, s) \in (0, \infty) \times \mathbb{C}_0^* \mid \tau \in (0, \theta), \tau|s| < \eta\}. \end{aligned}$$

The idea of the proof is to compare $\mathbf{H}_\tau(e^{\tau s})$ to $\mathbf{G}(s)$ for all $(\tau, s) \in V^*(\theta, \eta)$, where $\eta > 0$ and $\theta > 0$ are sufficiently small. We proceed in several steps.

Step 1: In this step we analyze the term $\mathbf{H}_\tau^1(e^{\tau s}) - \mathbf{G}_1(s)$. We claim that

$$(4.39) \quad \mathbf{G}_1(s) - \mathbf{H}_\tau^1(e^{\tau s}) = F_1(sI - A_1)^{-1}Q(\tau, s)B_1, \quad \forall (\tau, s) \in (0, \infty) \times \mathbb{C}_0^*,$$

where Q is a matrix-valued function defined on $(0, \infty) \times \mathbb{C}_0^*$ such that

$$(4.40) \quad \lim_{(\theta, \eta) \rightarrow 0} (\sup\{\|Q(\tau, s)\| \mid (\tau, s) \in V^*(\theta, \eta)\}) = 0.$$

Using the finite-dimensionality of X_1 , we obtain using Taylor expansions

$$(4.41) \quad \int_0^\tau T_1(t)B_1 dt = \tau(I + Q_1(\tau))B_1, \quad I - e^{-\tau s}T_1(\tau) = \tau(I + Q_2(\tau, s))(sI - A_1),$$

where

$$Q_1(\tau) := \sum_{k=2}^{\infty} \frac{\tau^{k-1}}{k!} A_1^{k-1}, \quad Q_2(\tau, s) := \sum_{k=2}^{\infty} \frac{\tau^{k-1}}{k!} (A_1 - sI)^{k-1}.$$

Note that Q_1 and Q_2 satisfy

$$(4.42) \qquad \lim_{\tau \rightarrow 0} Q_1(\tau) = 0, \qquad \lim_{(\tau, \tau s) \rightarrow 0} Q_2(\tau, s) = 0.$$

The first limit follows immediately from the definition of Q_1 , whilst the second limit follows from the inequality

$$\|Q_2(\tau, s)\| \leq e^{\tau(\|s\| + \|A_1\|)} - 1,$$

which is the result of a routine estimate. Using (4.41), a straightforward calculation shows that (4.39) holds with

$$(4.43) \quad Q(\tau, s) := \tau[(1 - e^{-\tau s})I - e^{-\tau s}Q_1(\tau) + Q_2(\tau, s)](sI - A_1)(I - e^{-\tau s}T_1(\tau))^{-1}.$$

Note that, by the spectral mapping theorem, $I - e^{-\tau s}T_1(\tau)$ is invertible for all $(\tau, s) \in (0, \infty) \times \mathbb{C}_0^*$, and so, $Q(\tau, s)$ is well-defined for all $(\tau, s) \in (0, \infty) \times \mathbb{C}_0^*$. From (4.41) we obtain

$$Q(\tau, s) = [(1 - e^{-\tau s})I - e^{-\tau s}Q_1(\tau) + Q_2(\tau, s)](I + Q_2(\tau, s))^{-1},$$

and it follows from (4.42) that Q satisfies (4.40).

Step 2: We show that

$$(4.44) \qquad \lim_{(\theta, \eta) \rightarrow 0} (\sup\{\|\mathbf{H}_\tau^2(e^{\tau s}) - \mathbf{G}_2(s)\| \mid (\tau, s) \in V(\theta, \eta)\}) = 0.$$

In order to estimate the difference $\mathbf{H}_\tau^2(e^{\tau s}) - \mathbf{G}_2(s)$, note that

$$\begin{aligned} \mathbf{H}_\tau^2(e^{\tau s}) - \mathbf{G}_2(s) &= F_2 \left[(e^{\tau s}I - T_2(\tau))^{-1}(T_2(\tau) - I)A_2^{-1} - (sI - A_2)^{-1} \right] B_2 \\ &= F_2 \left[(T_2(\tau) - I)A_2^{-1}(sI - A_2) - (e^{\tau s}I - T_2(\tau)) \right] \\ &\qquad \times (e^{\tau s}I - T_2(\tau))^{-1}(sI - A_2)^{-1}B_2. \end{aligned}$$

Setting $R_2(s) := (sI - A_2)^{-1}$, routine algebraic manipulations show that

$$(4.45) \qquad \mathbf{H}_\tau^2(e^{\tau s}) - \mathbf{G}_2(s) = P_1(\tau, s)P_2(\tau, s),$$

where

$$\begin{aligned} P_1(\tau, s) &:= F_2 \left(e^{-\tau s} \frac{T_2(\tau) - I}{\tau} A_2^{-1} - \frac{1 - e^{-\tau s}}{\tau s} I \right), \\ P_2(\tau, s) &:= \left(\frac{I - e^{-\tau s}T_2(\tau)}{\tau s} \right)^{-1} R_2(s)B_2. \end{aligned}$$

We first analyze $P_1(\tau, s)$. Since the term

$$e^{-\tau s} \frac{T_2(\tau) - I}{\tau} A_2^{-1} - \frac{1 - e^{-\tau s}}{\tau s} I$$

converges strongly to 0 as $(\tau, \tau s) \rightarrow 0$, the compactness of F_2 together with Lemma 2.1 shows that

$$(4.46) \qquad \lim_{(\tau, \tau s) \rightarrow 0} \|P_1(\tau, s)\| = 0.$$

Estimating $P_2(\tau, s)$ is considerably more difficult. In view of (4.45) and (4.46), to establish (4.44), it is sufficient to prove the following claim.

Claim: For any $p \in (0, \pi)$ there exist $\delta > 0$ and $k > 0$ such that

$$(4.47) \qquad \|P_2(\tau, s)\| \leq k < \infty, \quad \forall (\tau, s) \in V(\delta, p).$$

Proof. Since the semigroup $T_2(t)$ is exponentially stable, for fixed $s \in \mathbb{C}_0$ and $\tau > 0$ the spectral radius of $e^{-\tau s}T_2(\tau)$ is smaller than 1, and the Neumann series gives

$$(4.48) \quad (I - e^{-\tau s}T_2(\tau))^{-1} = \sum_{k=0}^{\infty} e^{-k\tau s}T_2(k\tau).$$

Since $T_2(t)$ is an exponentially stable analytic semigroup, there exist $\beta \in (0, \pi/2)$, $\omega > 0$ and $M > 0$ such that

$$\varrho(A_2) \supset \Xi := \{\xi \in \mathbb{C} \mid |\arg(\xi + \omega)| < \pi/2 + \beta\}$$

and

$$(4.49) \quad \|R_2(\xi)\| \leq M/|\xi|, \quad \forall \xi \in \Xi \setminus \{0\}.$$

If we note that $R_2(\xi)B_2 = (\xi R_2(\xi) - I)A_2^{-1}B_2$, (4.49) shows that for any $\varepsilon > 0$ there exists $N > 0$ such that

$$(4.50) \quad \|R_2(\xi)B_2\|_{\mathcal{B}(U,X)} \leq N, \quad \forall \xi \in \Xi.$$

Let $\psi \in (\pi/2, \pi/2 + \beta)$ and $r_0 \in (0, \omega)$. Define the contour

$$\Gamma = \{re^{-i\psi} \mid r > r_0\} \cup \{r_0e^{i\phi} \mid \phi \in [\psi, 2\pi - \psi]\} \cup \{re^{i\psi} \mid r > r_0\},$$

oriented bottom to top. Note that by construction $\Gamma \subset \Xi$ and $\sigma(A_2)$ is to the left of Γ ; see Figure 4.1.

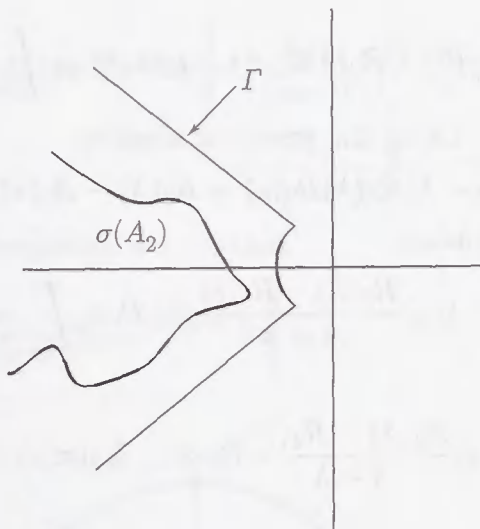


FIGURE 4.1. The contour Γ with $\sigma(A_2)$ to the left of Γ

Invoking (4.48) and using a standard relationship between $T_2(t)$ and $R_2(s)$ (see [14], Theorem 7.7, p. 30), we have

$$(4.51) \quad (I - e^{-\tau s}T_2(\tau))^{-1} = I + \frac{1}{2\pi i} \sum_{k=1}^{\infty} \int_{\Gamma} e^{-k\tau s} e^{k\tau \lambda} R_2(\lambda) d\lambda.$$

By (4.49),

$$\|e^{-k\tau s} e^{k\tau \lambda} R_2(\lambda)\| \leq \frac{M}{|\lambda|} |e^{k\tau(\lambda-s)}|, \quad \forall \lambda \in \Gamma, \forall s \in \mathbb{C}_0, \forall \tau > 0.$$

Hence we can use Lebesgue's dominated convergence theorem to interchange the order of summation and integration in (4.51). Evaluating the resulting infinite sum, we obtain

$$(4.52) \quad P_2(\tau, s) = \tau s R_2(s) B_2 + \frac{1}{2\pi i} \int_{\Gamma} \frac{\tau s e^{\tau(\lambda-s)}}{1 - e^{\tau(\lambda-s)}} R_2(\lambda) R_2(s) B_2 d\lambda.$$

Setting

$$(4.53) \quad f(\tau, \lambda, s) := \frac{\tau s e^{\tau(\lambda-s)}}{1 - e^{\tau(\lambda-s)}} + \frac{s}{\lambda - s} = \frac{\tau s}{e^{\tau(s-\lambda)} - 1} - \frac{s}{s - \lambda},$$

and letting $p \in (0, \pi)$ be fixed, but arbitrary, an elementary analysis (see the Appendix) shows that

$$(4.54) \quad \sup \left\{ \left| \frac{f(\tau, \lambda, s)}{\tau(s - \lambda)} \right| \mid (\tau, \lambda, s) \in (0, \infty) \times \Gamma \times \mathbb{C}_0 \text{ s.t. } \tau|\lambda| \leq p, \tau|s| \leq p \right\} < \infty.$$

Define $\delta := p/r_0$ and for $\tau \in (0, \delta)$ introduce

$$\Gamma_{\tau} := \{\lambda \in \Gamma \mid \tau|\lambda| \leq p\}.$$

Clearly, $\Gamma_{\tau} \neq \emptyset$ for all $\tau \in (0, \delta)$. We decompose the integral on the right-hand side of (4.52) as follows:

$$(4.55) \quad I(\tau, s) := \int_{\Gamma} \frac{\tau s e^{\tau(\lambda-s)}}{1 - e^{\tau(\lambda-s)}} R_2(\lambda) R_2(s) B_2 d\lambda = I_1(\tau, s) + I_2(\tau, s) R_2(s) B_2,$$

where

$$I_1(\tau, s) := \int_{\Gamma_{\tau}} \frac{\tau s e^{\tau(\lambda-s)}}{1 - e^{\tau(\lambda-s)}} R_2(\lambda) R_2(s) B_2 d\lambda, \quad I_2(\tau, s) := \int_{\Gamma \setminus \Gamma_{\tau}} \frac{\tau s e^{\tau(\lambda-s)}}{1 - e^{\tau(\lambda-s)}} R_2(\lambda) d\lambda.$$

We first analyze $I_1(\tau, s)$. Using the resolvent identity

$$(s - \lambda) R_2(\lambda) R_2(s) = R_2(\lambda) - R_2(s),$$

and invoking (4.53), we obtain

$$(4.56) \quad I_1(\tau, s) = \int_{\Gamma_{\tau}} f(\tau, \lambda, s) \frac{R_2(\lambda) - R_2(s)}{s - \lambda} B_2 d\lambda + \int_{\Gamma_{\tau}} \frac{s}{s - \lambda} R_2(\lambda) R_2(s) B_2 d\lambda.$$

Setting

$$I_{11}(\tau, s) := \int_{\Gamma_{\tau}} f(\tau, \lambda, s) \frac{R_2(\lambda) - R_2(s)}{s - \lambda} B_2 d\lambda, \quad I_{12}(\tau, s) := \int_{\Gamma_{\tau}} \frac{s}{s - \lambda} R_2(\lambda) d\lambda,$$

we have

$$(4.57) \quad I_1(\tau, s) = I_{11}(\tau, s) + I_{12}(\tau, s) R_2(s) B_2.$$

Invoking (4.50) and (4.54) shows that there exists $k_1 > 0$ such that for all $(\tau, s) \in V(\delta, p)$,

$$\begin{aligned} \|I_{11}(\tau, s)\| &\leq \int_{\Gamma_{\tau}} \frac{|f(\tau, \lambda, s)|}{|\lambda - s|} (\|R_2(\lambda) B_2\| + \|R_2(s) B_2\|) d|\lambda| \\ &\leq k_1 \tau \text{length}(\Gamma_{\tau}). \end{aligned}$$

Combining this with

$$\text{length}(\Gamma_{\tau}) \leq 2(\pi + 1)p/\tau$$

and setting $k_2 := 2k_1(\pi + 1)p$ yields

$$(4.58) \quad \|I_{11}(\tau, s)\| \leq k_2, \quad \forall (\tau, s) \in V(\delta, p).$$

To estimate $I_{12}(\tau, s)$, we define

$$\kappa(\tau) := 2p/\tau,$$

and for each $\tau \in (0, \delta)$ we embed Γ_τ into a closed contour

$$C_\tau := \{re^{-i\psi} \mid r \in [r_0, \kappa(\tau)]\} \cup \{r_0 e^{i\phi} \mid \phi \in [\psi, 2\pi - \psi]\} \\ \cup \{re^{i\psi} \mid r \in [r_0, \kappa(\tau)]\} \cup \{\kappa(\tau) e^{i\phi} \mid \phi \in [-\psi, \psi]\},$$

oriented clockwise (see Figure 4.2). Then

$$(4.59) \quad I_{12}(\tau, s) = \int_{C_\tau} \frac{s}{(s-\lambda)} R_2(\lambda) d\lambda - \int_{p/\tau}^{\kappa(\tau)} \frac{s}{s-re^{i\psi}} R_2(re^{i\psi}) e^{i\psi} dr \\ + \int_{p/\tau}^{\kappa(\tau)} \frac{s}{s-re^{-i\psi}} R_2(re^{-i\psi}) e^{-i\psi} dr + \int_{-\psi}^{\psi} \frac{is\kappa(\tau)e^{i\phi}}{s-\kappa(\tau)e^{i\phi}} R_2(\kappa(\tau)e^{i\phi}) d\phi.$$

By construction, for a given $\tau \in (0, \delta)$, any $s \in \mathbb{C}_0 \cap \mathbb{B}(0, p/\tau)$ is inside the contour C_τ and so Cauchy's integral formula combined with (4.49) shows that there exists $k_3 > 0$ such that

$$(4.60) \quad \left\| \int_{C_\tau} \frac{s}{(s-\lambda)} R_2(\lambda) d\lambda \right\| = 2\pi \|sR_2(s)\| \leq k_3, \quad \forall (\tau, s) \in V(\delta, p).$$

Moreover, since $|s - re^{i\psi}| > r|\cos \psi|$ for $s \in \mathbb{C}_0$, it follows from (4.49) that there exists $k_4 > 0$ such that

$$\left\| \int_{p/\tau}^{\kappa(\tau)} \frac{s}{s-re^{i\psi}} R_2(re^{i\psi}) e^{i\psi} dr \right\| \leq \frac{k_4|s|}{|\cos \psi|} \int_{p/\tau}^{\infty} \frac{dr}{r^2} \\ = \frac{k_4}{p|\cos \psi|} \tau |s|, \quad \forall (\tau, s) \in V(\delta, p).$$

As an immediate consequence we see that

$$(4.61) \quad \left\| \int_{p/\tau}^{\kappa(\tau)} \frac{s}{s-re^{i\psi}} R_2(re^{i\psi}) e^{i\psi} dr \right\| \leq k_5, \quad \forall (\tau, s) \in V(\delta, p),$$

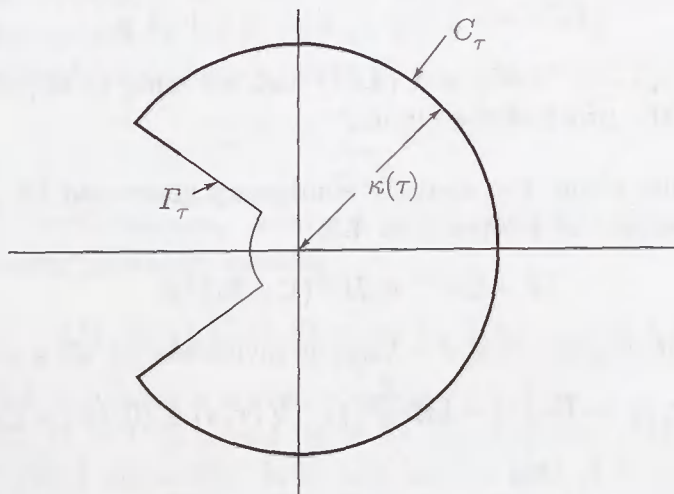


FIGURE 4.2. The contour Γ_τ embedded in a closed contour C_τ

for some suitable $k_5 > 0$. Similarly, there exists $k_6 > 0$ such that

$$(4.62) \qquad \left\| \int_{p/\tau}^{\kappa(\tau)} \frac{s}{s - re^{-i\psi}} R_2(re^{-i\psi}) e^{-i\psi} dr \right\| \leq k_6, \quad \forall (\tau, s) \in V(\delta, p).$$

Trivially, since $\kappa(\tau) = 2p/\tau$,

$$|s - \kappa(\tau)e^{i\phi}| \geq p/\tau, \quad \forall (\tau, s) \in V(\delta, p), \forall \phi \in [-\psi, \psi],$$

and hence, using (4.49), we see that there exists $k_7 > 0$ such that

$$(4.63) \qquad \left\| \int_{-\psi}^{\psi} \frac{is\kappa(\tau)e^{i\phi}}{s - \kappa(\tau)e^{i\phi}} R_2(\kappa(\tau)e^{i\phi}) d\phi \right\| \leq \frac{k_7\tau|s|}{p} \int_{-\psi}^{\psi} d\phi = \frac{2k_7\psi\tau|s|}{p} \\ \leq 2k_7\psi, \quad \forall (\tau, s) \in V(\delta, p).$$

Combining (4.59)–(4.63) shows that there exists $k_8 > 0$ such that

$$(4.64) \qquad \|I_{12}(\tau, s)\| \leq k_8, \quad \forall (\tau, s) \in V(\delta, p).$$

By (4.50) we may conclude that $I_{12}(\tau, s)R_2(s)B_2$ is uniformly bounded on $V(\delta, p)$, and therefore, by (4.57) and (4.58), we obtain that there exists $k_9 > 0$ such that

$$(4.65) \qquad \|I_1(\tau, s)\| \leq k_9, \quad \forall (\tau, s) \in V(\delta, p).$$

To estimate $I_2(\tau, s)$, first note that there exists $k_{10} > 0$ such that

$$|1 - e^{\tau(\lambda-s)}| \geq k_{10} > 0, \quad \forall (\tau, \lambda, s) \in (0, \infty) \times \Gamma \times \mathbb{C}_0 \text{ s.t. } \lambda \in \Gamma \setminus \Gamma_\tau.$$

Combining this with an application of (4.49), we obtain

$$(4.66) \qquad \|I_2(\tau, s)\| \leq \frac{2k_{11}p}{k_{10}} \int_{p/\tau}^{\infty} \frac{e^{\tau r \cos \psi}}{r} dr, \quad \forall (\tau, s) \in V(\delta, p),$$

for some suitable constant $k_{11} > 0$. Since $\cos \psi < 0$,

$$\int_{p/\tau}^{\infty} \frac{e^{\tau r \cos \psi}}{r} dr \leq \frac{\tau}{p\tau|\cos \psi|} e^{p \cos \psi} \leq \frac{e^{p \cos \psi}}{p|\cos \psi|} < \infty.$$

Together with (4.66) this shows that there exists $k_{12} > 0$ such that

$$(4.67) \qquad \|I_2(\tau, s)\| \leq k_{12}, \quad \forall (\tau, s) \in V(\delta, p).$$

Combining (4.52), (4.55), (4.65) and (4.67) and invoking (4.50) shows that (4.47) holds, completing the proof of the claim. □

Step 3: By assumption, the analytic semigroup generated by A_{BF} is exponentially stable. Therefore, by Proposition 4.5,

$$(4.68) \qquad (I - \mathbf{G})^{-1} \in H^\infty(\mathbb{C}_0, \mathcal{B}(U)),$$

which, in particular, implies that $I - \mathbf{G}(s)$ is invertible for all $s \in \mathbb{C}_0^*$. Setting

$$(4.69) \qquad D(\tau, s) := \mathbf{G}_2(s) - \mathbf{H}_\tau^2(e^{s\tau}), \quad \forall (\tau, s) \in (0, \infty) \times \mathbb{C}_0,$$

and

$$(4.70) \qquad \tilde{Q}(\tau, s) := \begin{pmatrix} Q(\tau, s) & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{B}(X_{-1}, X), \quad \forall (\tau, s) \in (0, \infty) \times \mathbb{C}_0^*,$$

with Q given by (4.43), we obtain from (4.38) and (4.39) for all $(\tau, s) \in (0, \infty) \times \mathbb{C}_0^*$ that

$$\begin{aligned} I - \mathbf{H}_\tau(e^{\tau s}) &= I - \mathbf{G}_1(s) - \mathbf{G}_2(s) + F_1(sI - A_1)^{-1}Q(\tau, s)B_1 + D(\tau, s) \\ &= (I - \mathbf{G}(s))[I + (I - \mathbf{G}(s))^{-1}F(sI - A)^{-1}\tilde{Q}(s, \tau)B \\ &\quad + (I - \mathbf{G}(s))^{-1}D(\tau, s)]. \end{aligned} \quad (4.71)$$

A straightforward calculation combined with an application of statements (2) and (3) of Lemma 4.1 shows that

$$\begin{aligned} (I - \mathbf{G}(s))^{-1}F(sI - A)^{-1}x &= F(sI - A - BF)^{-1}x \\ &= F(sI - A_{BF})^{-1}x, \quad \forall s \in \mathbb{C}_0^*, \forall x \in X. \end{aligned} \quad (4.72)$$

Moreover, by (4.70),

$$(\tilde{Q}(\tau, s)B)U \subset X, \quad \forall (\tau, s) \in (0, \infty) \times \mathbb{C}_0^*. \quad (4.73)$$

Combining (4.71)–(4.73) shows that for all $(\tau, s) \in (0, \infty) \times \mathbb{C}_0^*$,

$$I - \mathbf{H}_\tau(e^{\tau s}) = (I - \mathbf{G}(s))[I + F(sI - A_{BF})^{-1}\tilde{Q}(\tau, s)B + (I - \mathbf{G}(s))^{-1}D(\tau, s)]. \quad (4.74)$$

By assumption, the analytic semigroup generated by A_{BF} is exponentially stable. Consequently, the function $s \mapsto (sI - A_{BF})^{-1}$ is in $H^\infty(\mathbb{C}_0, \mathcal{B}(X))$, and thus we may conclude from (4.40) and (4.70) that

$$\lim_{(\theta, \eta) \rightarrow 0} (\sup\{\|F(sI - A_{BF})^{-1}\tilde{Q}(\tau, s)B\| \mid (\tau, s) \in V^*(\theta, \eta)\}) = 0. \quad (4.75)$$

From (4.44) and (4.69) combined with (4.68), we obtain

$$\lim_{(\theta, \eta) \rightarrow 0} (\sup\{\|(I - \mathbf{G}(s))^{-1}D(\tau, s)\| \mid (\tau, s) \in V(\theta, \eta)\}) = 0. \quad (4.76)$$

Finally, combining (4.74)–(4.76) shows that there exist $\theta > 0$ and $\eta > 0$ such that (4.26) holds. \square

5. EXAMPLES

In this section we give examples which illustrate Theorem 4.8. We consider a general parabolic partial differential equation with Dirichlet control. For two different types of stabilizing feedbacks $u = Fx$ found in the literature on control of partial differential equations, we show that the associated sampled-data feedback (1.2) stabilizes the system for small enough $\tau > 0$.

Let Ω be a convex bounded open set in \mathbb{R}^n , $n \geq 2$, with C^∞ boundary Γ . Let

$$L(\xi, \partial) = \sum_{|\alpha| \leq 2} a_\alpha(\xi) \partial^\alpha$$

with smooth real coefficients a_α , where ∂ is the spatial derivative operator. We consider the following parabolic system:

$$(5.1) \quad \frac{\partial x}{\partial t}(t, \xi) = -L(\xi, \partial)x(t, \xi) \text{ in } (0, \infty) \times \Omega, \quad x(t, \zeta) = u(t, \zeta) \text{ in } (0, \infty) \times \Gamma.$$

An example of such a system is a heat equation in a disk.

Let $X = L^2(\Omega)$, $U = L^2(\Gamma)$, and let A be the operator $-L(\xi, \partial)$ with domain $\mathcal{D}(A) = \{x \in H^2(\Omega) \mid x|_\Gamma = 0\}$. It is well known that A is the generator of an analytic semigroup; see for instance [4]. Define the Dirichlet map D by

$$x = Dy \text{ if } L(\xi, \partial)x = 0 \text{ in } \Omega \text{ and } x = y \text{ on } \Gamma.$$

Feedback stabilization of (5.1) is studied in Lasiecka and Triggiani [10], [11], [12]. In [11] it is shown that (5.1) can be written in the form $\dot{x} = Ax + Bu$ with $B = -AD$. From standard elliptic theory, $D \in \mathcal{B}(L^2(\Gamma), H^{1/2}(\Omega))$ (see for instance [13]); so we see that $B \in \mathcal{B}(U, X_{-1})$.

We now describe two examples of compact feedbacks that exponentially stabilize (5.1).

Example 1. We first consider the rank m feedback from [10] of the form

$$Fx(t, \cdot) = \sum_{j=1}^m \langle x(t, \cdot), w_j(\cdot) \rangle_{L^2(\Omega)} g_j(\cdot),$$

where $w_j \in L^2(\Omega)$ and $g_j \in L^2(\Gamma)$. Theorem 1.2 from [10] gives conditions on w_j and g_j that guarantee that $A + BF$ generates an exponentially stable semigroup. Since F is obviously compact, we can apply Theorem 4.8 to conclude that (1.2) exponentially stabilizes (5.1) for small enough $\tau > 0$.

Example 2. Here the stabilizing feedback arises from the solution of a linear quadratic optimal control problem. Consider the cost functional

$$J(u, x) = \int_0^\infty (\|u(t)\|_{L^2(\Gamma)}^2 + \|x(t)\|_{L^2(\Omega)}^2) dt,$$

where x is the solution of (5.1) corresponding to the control function u . It is shown in [11] that the optimal control is given by the feedback

$$(5.2) \quad u = \frac{\partial}{\partial \nu_{A^*}} Px,$$

where $\partial/\partial \nu_{A^*}$ is the co-normal derivative with respect to A^* and $P \in \mathcal{B}(L^2(\Omega))$ is a nonnegative selfadjoint operator which is the unique solution of the algebraic Riccati equation

$$\langle x, y \rangle_{L^2(\Omega)} + \langle Px, Ay \rangle_{L^2(\Omega)} + \langle Ax, Py \rangle_{L^2(\Omega)} = \left\langle \frac{\partial}{\partial \nu_{A^*}} Px, \frac{\partial}{\partial \nu_{A^*}} Py \right\rangle_{L^2(\Gamma)}.$$

It is shown in Theorem 2.8 in [11] that the control (5.2) can be written as $u = Fx$, where $F \in \mathcal{B}(X, U)$. Furthermore, it is shown in the proof of Theorem 2.10 in [11] that BF can be written in the form AK , where $K \in \mathcal{B}(X)$ is compact. Hence we can apply Theorem 4.8 to $\dot{x} = Ax + Av$ with feedback $v = Kx$, to conclude that (1.2) exponentially stabilizes (5.1) for small enough $\tau > 0$. In [12], Galerkin approximations of the feedback (5.2) were developed and shown to be exponentially stabilizing for sufficiently small mesh parameters. These Galerkin approximations produce finite-rank feedback, and therefore Theorem 4.8 can again be applied.

6. DYNAMIC SAMPLED-DATA FEEDBACK

In this section we apply the results of the previous sections to dynamic sampled-data feedback. For simplicity and the sake of brevity, we restrict ourselves here to systems with bounded control and bounded observation. The system to be controlled, the plant, is formally given by

$$(6.1a) \quad \dot{x}(t) = Ax(t) + Bu(t), \quad t \geq 0; \quad x(0) = x^0 \in X,$$

$$(6.1b) \quad y(t) = Cx(t),$$

where $A : D(A) \subset X \rightarrow X$ generates a strongly continuous semigroup $T(t)$, $B \in \mathcal{B}(U, X)$, $C \in \mathcal{B}(X, Y)$, and X , U and Y are Hilbert spaces.

Consider a continuous-time controller of the form

$$(6.2a) \quad \dot{z}(t) = A_c z(t) + B_c v(t), \quad t \geq 0; \quad z(0) = z^0 \in X_c,$$

$$(6.2b) \quad w(t) = C_c z(t) + D_c v(t),$$

where $A_c : D(A_c) \subset X_c \rightarrow X_c$ generates a strongly continuous semigroup $T_c(t)$, $B_c \in \mathcal{B}(Y, X_c)$, $C_c \in \mathcal{B}(X_c, U)$, $D_c \in \mathcal{B}(Y, U)$ and X_c is a Hilbert space. Using the output (6.1b) of (6.1) as the input for (6.2) and the output (6.2b) of (6.2) as the input for (6.1), i.e.,

$$(6.3) \quad v = y \quad \text{and} \quad u = w,$$

we obtain the feedback interconnection of (6.1) and (6.2). It is convenient to define

$$\tilde{A} := \begin{pmatrix} A & 0 \\ 0 & A_c \end{pmatrix}, \quad \tilde{B} := \begin{pmatrix} B & 0 \\ 0 & B_c \end{pmatrix}, \quad \tilde{C} := \begin{pmatrix} C & 0 \\ 0 & C_c \end{pmatrix}, \quad \tilde{K} := \begin{pmatrix} D_c & I_U \\ I_Y & 0 \end{pmatrix}.$$

Clearly, \tilde{A} generates the strongly continuous semigroup

$$\tilde{T}(t) := \begin{pmatrix} T(t) & 0 \\ 0 & T_c(t) \end{pmatrix}.$$

We say that the continuous-time feedback system given by (6.1)–(6.3) is *exponentially stable* if the strongly continuous semigroup generated by

$$\tilde{A} + \tilde{B}\tilde{K}\tilde{C} = \begin{pmatrix} A + BD_cC & BC_c \\ B_cC & A_c \end{pmatrix}$$

is exponentially stable.

Let $\tau > 0$ and consider the following discretization of (6.2), which is obtained by applying the standard hold and sampling operations to (6.2):

$$(6.4a) \quad z_{k+1} = T_c(\tau)z_k + \int_0^\tau T_c(s)B_c v_k ds, \quad k \in \mathbb{N}_0; \quad z_0 = z^0 \in X_c,$$

$$(6.4b) \quad w_k = C_c z_k + D_c v_k.$$

We use the discrete-time system (6.4) to control the continuous-time system (6.1) by sampled-data feedback, i.e., we consider the feedback interconnection of (6.1) and (6.4) given by

$$v_k = y(k\tau) \quad \text{and} \quad u(k\tau + t) = w_k, \quad t \in [0, \tau); \quad k \in \mathbb{N}_0.$$

Via an application of the variation of parameters formula to (6.1a), this leads to the following sampled-data feedback equations:

$$(6.5a)$$

$$x(k\tau + t) = T(t)x(k\tau) + \int_0^t T(s)B(D_c C x(k\tau) + C_c z_k) ds, \quad t \in [0, \tau); \quad x(0) = x^0,$$

$$(6.5b)$$

$$z_{k+1} = T_c(\tau)z_k + \int_0^\tau T_c(s)B_c C x(k\tau) ds, \quad k \in \mathbb{N}_0; \quad z_0 = z^0.$$

The sampled-data feedback system (6.5) is called *exponentially stable* if there exist $N \geq 1$ and $\nu > 0$ such that for all initial conditions $(x^0, z^0) \in X \times X_c$,

$$\|(x(k\tau + t), z_k)\| \leq N e^{-\nu(k\tau + t)} \|(x^0, z^0)\|, \quad \forall t \in [0, \tau), \quad \forall k \in \mathbb{N}_0.$$

By a straightforward calculation it follows from (6.4a) that

$$\begin{pmatrix} x((k+1)\tau) \\ z_{k+1} \end{pmatrix} = \tilde{\Delta}_\tau \begin{pmatrix} x(k\tau) \\ z_k \end{pmatrix}, \quad k \in \mathbb{N}_0,$$

where

$$\tilde{\Delta}_\tau : X \times X_c \rightarrow X \times X_c, \quad \tilde{x} \mapsto \tilde{T}(\tau)\tilde{x} + \int_0^\tau \tilde{T}(s)\tilde{B}\tilde{K}\tilde{C}\tilde{x} \, ds.$$

The proof of the following lemma is a routine exercise and is therefore left to the reader.

Lemma 6.1. *For any $\tau > 0$, the sampled-data feedback system (6.5) is exponentially stable if and only if $\tilde{\Delta}_\tau$ is power stable.*

We are now in position to formulate the main result of this section.

Theorem 6.2. *Assume that the operator $\tilde{K}\tilde{C}$ is compact. If the continuous-time feedback system (6.1)–(6.3) is exponentially stable (i.e., the strongly continuous semigroup generated by $\tilde{A} + \tilde{B}\tilde{K}\tilde{C}$ is exponentially stable), then there exists $\tau^* > 0$ such that for every $\tau \in (0, \tau^*)$ the sampled-data feedback system (6.5) is exponentially stable.*

Proof. Introducing the operator

$$\tilde{S}_\tau : U \times Y \rightarrow X \times X_c, \quad \tilde{u} \mapsto \int_0^\tau \tilde{T}(s)\tilde{B}\tilde{u} \, ds,$$

we may write $\tilde{\Delta}_\tau = \tilde{T}(\tau) + \tilde{S}_\tau\tilde{K}\tilde{C}$, and we see that $\tilde{\Delta}$ is of the form (2.10). By the compactness of $\tilde{K}\tilde{C}$ and the exponential stability of the semigroup generated by $\tilde{A} + \tilde{B}\tilde{K}\tilde{C}$, it follows from Theorem 3.1 and Lemma 2.3 that there exists $\tau^* > 0$ such that for every $\tau \in (0, \tau^*)$ the operator $\tilde{\Delta}_\tau$ is power stable. An application of Lemma 6.1 yields the claim. \square

Trivially, the compactness assumption on $\tilde{K}\tilde{C}$ is satisfied if the observation operators C and C_c are compact. In particular, compactness of $\tilde{K}\tilde{C}$ is guaranteed if U and Y are finite-dimensional.

Under suitable assumptions, Theorem 4.8 can be used to extend Theorem 6.2 to systems and controllers with unbounded control operators B and B_c , respectively, provided the semigroups $T(t)$ and $T_c(t)$ are analytic. For the sake of brevity we omit the details.

APPENDIX: DERIVATION OF (4.54)

Let $p \in (0, \pi)$ and consider the meromorphic function g defined on \mathbb{C} by

$$g(\xi) := \frac{1 + \xi - e^\xi}{\xi(e^\xi - 1)}.$$

Clearly, for any $\varepsilon \in (0, 2p)$, the function g is bounded on $\mathbb{B}(0, 2p) \setminus \mathbb{B}(0, \varepsilon)$. Moreover, invoking the Taylor expansion of e^ξ or using L'Hôpital's rule, we have

$$\lim_{\xi \rightarrow 0} g(\xi) = -1/2.$$

Therefore, we may conclude that

$$(A.1) \quad \sup\{|g(\xi)| \mid \xi \in \mathbb{B}(0, 2p), \xi \neq 0\} < \infty.$$

By a straightforward calculation it follows from (4.53) that

$$\frac{|f(\tau, \lambda, s)|}{|\tau(s - \lambda)|} = \frac{|s|}{|s - \lambda|} |g(\tau(s - \lambda))|.$$

Noting that

$$\sup\{|s|/|s - \lambda| \mid (\lambda, s) \in \Gamma \times \mathbb{C}_0\} < \infty,$$

we see from (A.1) that

$$\sup\left\{\left|\frac{f(\tau, \lambda, s)}{\tau(s - \lambda)}\right| \mid (\tau, \lambda, s) \in (0, \infty) \times \Gamma \times \mathbb{C}_0 \text{ s.t. } \tau|\lambda| \leq p, \tau|s| \leq p\right\} < \infty,$$

which is (4.54).

REFERENCES

- [1] T. Chen and B. Francis, Input-output stability of sampled-data systems, *IEEE Trans. Automat. Control* **36** (1991), pp. 50-58. MR **91j**:93094
- [2] T. Chen and B. Francis, *Optimal Sampled-Data Control Systems*, Springer, London, 1995. MR **99a**:93001
- [3] R. F. Curtain and H. J. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*, Springer-Verlag, New York, 1995. MR **96i**:93001
- [4] N. Dunford and J. Schwartz, *Linear Operators, Part II*, John Wiley and Sons, New York, 1963. MR **32**:6181
- [5] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics, No. 194, Springer-Verlag, New York, 2000. MR **2000i**:47075
- [6] G. F. Franklin, J. D. Powell and M. Workman, *Digital Control of Dynamic Systems*, 3rd edition, Addison Wesley, Menlo Park, 1998.
- [7] R. Kalman, B. Ho and K. Narendra, Controllability of linear dynamical systems, *Contributions to Differential Equations* **1** (1963), pp. 188-213. MR **27**:5012
- [8] T. Kato, *Perturbation Theory for Linear Operators*, 2nd edition, Springer-Verlag, Berlin, 1980. MR **96a**:47025
- [9] H. Logemann, Stability and stabilizability of linear infinite-dimensional discrete-time systems, *IMA J. of Mathematical Control & Information* **9** (1992), pp. 255-263. MR **94c**:93094
- [10] I. Lasiecka and R. Triggiani, Stabilization and structural assignment of Dirichlet boundary feedback parabolic equations, *SIAM J. Control* **21** (1983), pp. 766-803. MR **85i**:93028
- [11] I. Lasiecka and R. Triggiani, The regulator problem for parabolic equations with Dirichlet boundary control, Part I: Riccati's feedback synthesis and regularity of optimal solution, *Appl. Math. Optim.* **16** (1987), pp. 147-168. MR **88g**:93063a
- [12] I. Lasiecka and R. Triggiani, The regulator problem for parabolic equations with Dirichlet boundary control, Part II: Galerkin approximation, *Appl. Math. Optim.* **16** (1987), pp. 187-216. MR **88g**:93063b
- [13] J.-L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications*, Vols. 1 and 2, Springer, Berlin, 1972. MR **50**:2670; MR **50**:2671
- [14] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences **44**, Springer-Verlag, New York, 1983. MR **85g**:47061
- [15] R. Rebarber and S. Townley, Stabilization of distributed parameter systems by piecewise polynomial control, *IEEE Trans. Automat. Control* **42** (1997), pp. 1254-1257. MR **98j**:93073
- [16] R. Rebarber and S. Townley, Generalized sampled data feedback control of distributed parameter systems, *Systems & Control Letters* **34** (1998), pp. 229-240. MR **99f**:93081
- [17] R. Rebarber and S. Townley, Non-robustness of closed-loop stability for infinite-dimensional systems under sample and hold, *IEEE Trans. Automat. Control* **47** (2002), pp. 1381-1385.
- [18] I. G. Rosen and C. Wang, On stabilizability and sampling for infinite-dimensional systems, *IEEE Trans. Automat. Control* **37** (1992), pp. 1653-1656. MR **93g**:93077
- [19] E. D. Sontag, *Mathematical Control Theory*, 2nd edition, Springer-Verlag, New York, 1998. MR **99k**:93001
- [20] T. J. Tarn, J. R. Zavgren and X. M. Zeng, Stabilization of infinite-dimensional systems with periodic gains and sampled output, *Automatica*, **24** (1988), pp. 95-99. MR **89d**:93087

- [21] G. Weiss, Admissibility of unbounded control operators, *SIAM J. Control & Optim.* **27** (1989), pp. 527-545. MR **90c**:93060
- [22] Y. Yamamoto, A function space approach to sampled data control systems and tracking problems, *IEEE Trans. Automat. Control* **39** (1994), pp. 703-713. MR **95b**:93116
- [23] J. Zabczyk. On decomposition of generators, *SIAM J. Control & Optim.* **16** (1978), pp. 523-534 (erratum in *SIAM J. Control & Optim.* **18** (1980), p. 325). MR **58**:23757; MR **81c**:47044

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF BATH, BATH BA2 7AY, UNITED KINGDOM

E-mail address: hl@maths.bath.ac.uk

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NEBRASKA-LINCOLN, LINCOLN, NEBRASKA 68588-0323

E-mail address: rrebarbe@math.unl.edu

SCHOOL OF MATHEMATICAL SCIENCES, UNIVERSITY OF EXETER, EXETER, EX4 4QE, UNITED KINGDOM

E-mail address: townley@maths.ex.ac.uk

APPROXIMATIONS FOR GABOR AND WAVELET FRAMES

DEGUANG HAN

ABSTRACT. Let ψ be a frame vector under the action of a collection of unitary operators \mathcal{U} . Motivated by the recent work of Frank, Paulsen and Tiballi and some application aspects of Gabor and wavelet frames, we consider the existence and uniqueness of the best approximation by normalized tight frame vectors. We prove that for any frame induced by a projective unitary representation for a countable discrete group, the best normalized tight frame (NTF) approximation exists and is unique. Therefore it applies to Gabor frames (including Gabor frames for subspaces) and frames induced by translation groups. Similar results hold for semi-orthogonal wavelet frames.

1. INTRODUCTION

A frame for a separable Hilbert space \mathcal{H} is a sequence $\{x_n\}$ of \mathcal{H} such that there exist $A, B > 0$ with the property that

$$A\|x\|^2 \leq \sum_n |\langle x, x_n \rangle|^2 \leq B\|x\|^2$$

holds for all $x \in \mathcal{H}$. The optimal constants (maximal for A and minimal for B) are called frame bounds. When $A = B = 1$, $\{x_n\}$ is called a normalized frame. Frames are generalizations of Riesz bases in the sense that a frame allows linear dependence among its elements. This redundancy property can lead to more freedom when constructing atoms, i.e., frame elements for specific types of expansions (cf. [BHW], [BT], [Dau]). From the geometric point of view, a frame for a Hilbert space \mathcal{H} is a compression (under an orthogonal projection) of a Riesz basis for a larger Hilbert space \mathcal{K} , while normalized tight frames are compressions of orthonormal bases (cf. [HL]).

Let $\{x_n\}$ be a frame for \mathcal{H} . Then S defined by

$$Sx = \sum_n \langle x, x_n \rangle x_n, \quad x \in \mathcal{H},$$

is a positive invertible bounded linear operator on \mathcal{H} , which is called the *frame operator* for $\{x_n\}$. A direct calculation yields

$$x = \sum_n \langle x, S^{-1/2} x_n \rangle S^{-1/2} x_n = \sum_n \langle x, S^{-1} x_n \rangle x_n \quad x \in \mathcal{H},$$

Received by the editors February 19, 2002.

2000 *Mathematics Subject Classification.* Primary 42C15, 46C05, 47B10.

Key words and phrases. Hilbert spaces, frames, unitary systems, approximation, Gabor family and Gabor frames, wavelet frames.

which implies that $\{S^{-1/2}x_n\}$ is a normalized tight frame. The frame $\{S^{-1}x_n\}$ is called the *canonical (or standard) dual* of $\{x_n\}$. In general, a frame $\{y_n\}$ is called a *dual* for $\{x_n\}$ if

$$x = \sum_n \langle x, y_n \rangle x_n$$

holds for every $x \in \mathcal{H}$. Two frames $\{x_n\}$ and $\{y_n\}$ for \mathcal{H} are said to be *similar* if there is a bounded invertible operator T on \mathcal{H} such that $Tx_n = y_n$ for all n . If two normalized tight frames are similar by an invertible operator T , then T must be a unitary operator [HL].

Motivated by the Gram-Schmidt process for a linear independent set $\{x_1, \dots, x_m\}$ and some quantum chemistry problems (cf. [Lo], [AEG1], [AEG2] and [GL]), Frank, Paulsen and Tiballi in [FPT] investigated the *symmetric approximations* of frames by normalized tight ones. Let $\{x_n\}$ be a frame for \mathcal{H} . A normalized tight frame $\{y_n\}$ for H is said to be a *symmetric approximation* of $\{x_n\}$ if it is *similar* to $\{x_n\}$ and the inequality

$$\sum_n \|z_n - x_n\|^2 \geq \sum_n \|y_n - x_n\|^2$$

is valid for all normalized tight frames $\{z_n\}$ of H that are similar to $\{x_n\}$. One of the main results in [FPT] states that the symmetric approximation for a frame $\{x_n\}$ is $\{S^{-1/2}x_n\}$, where S is the frame operator.

Another approximation was introduced by Balan in [Ba] by using the *closeness bound*. A frame $\{y_n\}$ is said to be *close* to a frame $\{x_n\}$ if there is a number $\lambda \geq 0$ such that

$$\left\| \sum_n c_n (y_n - x_n) \right\| \leq \lambda \left\| \sum_n c_n x_n \right\|$$

for all $c = \{c_n\} \in l^2(\mathbb{N})$. The infimum of such λ is called the *closeness bound* of the frame $\{y_n\}$ to the frame $\{x_n\}$ and is denoted by $c(\{x_n\}, \{y_n\})$. The closeness relation is not an equivalence relation since it is not reflexive in general. R. Balan defined the distance, $d_b(\{x_n\}, \{y_n\})$, between $\{y_n\}$ and $\{x_n\}$ by $\log(d^0(\{x_n\}, \{y_n\}))$, where

$$d^0(\{x_n\}, \{y_n\}) = \max(c(\{x_n\}, \{y_n\}), c(\{y_n\}, \{x_n\})).$$

It turns out that (see Theorem 2.4 in [Ba]) $d_b(\{x_n\}, \{y_n\}) < \infty$ if and only if $\{x_n\}$ and $\{y_n\}$ are similar. In this case we in fact have

$$d_b(\{x_n\}, \{y_n\}) = \max(\|I - Q\|, \|I - Q^{-1}\|),$$

where Q is the invertible operator that induces the similarity between the two frames. Therefore the distance between inequivalent frames is always infinity. Balan is also able to compute the minimal distance between a given frame and the tight frames.

In applications the most important and practical frames are the ones that are generated by a single vector in a Hilbert space under the action of a suitable collection of unitary operators. Wavelet frames and Gabor frames are typical examples. A *unitary system* \mathcal{U} is a countable set of unitary operators acting on a separable Hilbert space \mathcal{H} that contains the identity operator. We say that a vector $\psi \in \mathcal{H}$ is a *complete frame vector* (resp. *complete normalized tight frame vector*) for \mathcal{U} if $\mathcal{U}\psi := \{U\psi : U \in \mathcal{U}\}$ is a frame (resp. normalized tight frame) for \mathcal{H} . When $\mathcal{U}\psi$ is an orthonormal basis for \mathcal{H} , ψ is called a *complete wandering vector* for \mathcal{U} . We remark that all these frames are special *uniform frames* (a uniform frame

is a frame with the property that all the frame elements have the same norm). For an extensive discussion of uniform frames (especially uniform tight frames in finite-dimensional spaces) we refer to the recent work [BF] and [CK].

Two frame vectors ψ and η for a unitary system \mathcal{U} are said to be *similar* (denoted by $\psi \sim \eta$) when the two frames $\mathcal{U}\psi$ and $\mathcal{U}\eta$ are similar. Write $\mathcal{T}(\mathcal{U})$ for the set of all the complete normalized tight frame vectors for \mathcal{U} . Given a frame vector ψ for a unitary system \mathcal{U} , it is natural to ask which normalized tight frame vector is the closest to ψ . We note that in this setting neither the symmetric approximation nor the Balan's distance makes sense for this problem since (a) for any two different vectors ξ and η , $\sum_{U \in \mathcal{U}} \|U\xi - U\eta\|^2 = \infty$ if \mathcal{U} is an infinite set and (b) $\mathcal{U}\xi$ and $\mathcal{U}\psi$ are not similar in general. We say that $\xi \in \mathcal{T}(\mathcal{U})$ is a *best normalized tight frame (NTF) approximation* for a given complete frame vector ψ for \mathcal{U} if

$$\|\xi - \psi\| = \text{dist}(\xi, \mathcal{T}(\mathcal{U})) := \inf\{\|\eta - \psi\| : \eta \in \mathcal{T}(\mathcal{U})\}.$$

For a given unitary system \mathcal{U} and a complete frame vector ψ , we might expect that there always exists a best NTF approximation for ψ . However, this is false, since there exist unitary systems \mathcal{U} that admit complete frame vectors but do not admit any normalized tight frame vector. (For instance, let U be the unitary operator on \mathbb{R}^2 that is the rotation operator by $\pi/4$, and let $\mathcal{U} = \{I, U\}$. Then \mathcal{U} admits complete frame vectors but does not admit tight ones. One can easily refine this to the infinite-dimensional complex Hilbert space case.) Therefore it is natural to ask the following:

Question 1. Suppose that a unitary system \mathcal{U} admits a complete normalized tight frame vector. When do we have a best NTF approximation for each complete frame vector? Is the NTF approximation always unique?

In the case of Gabor frames this question has been addressed previously by Daubechies, Frank, Feichtinger, Paulsen and Tiballi, etc. The purpose of this paper is to answer this question for some important frames including Gabor frames and wavelet frames.

We recall some definitions and notation about Gabor and wavelet frames. Let Λ be a full-rank lattice in $\mathbb{R}^d \times \mathbb{R}^d$, and let $g(x) \in L^2(\mathbb{R}^d)$. The *Gabor family* associated with Λ and g is the collection:

$$\mathcal{G}(\Lambda, g) = \{e^{2\pi i \langle \ell_1, x \rangle} g(x - \ell_2) : (\ell_1, \ell_2) \in \Lambda\}.$$

Such a family was first introduced by Gabor [Ga] in 1946 for the purpose of signal processing and is still both theoretically appealing and successfully used in applications. For some recent developments we refer to the book [FS] by Feichtinger and Strohmer, and a survey paper [Ca2] by Casazza. When $\mathcal{G}(\Lambda, g)$ is a frame for $L^2(\mathbb{R}^d)$, we call g a *Gabor frame generator*.

Let M be a real expansive matrix (i.e., all the eigenvalues of M are required to have absolute values greater than 1). An (M -dilation) *wavelet frame* (resp. *normalized tight wavelet frame*) is a single function $\psi \in L^2(\mathbb{R}^d)$ with the property that

$$\psi_{m,\ell}(x) := |\det M|^{\frac{m}{2}} \psi(M^m x - \ell) : m \in \mathbb{Z}, \ell \in \mathbb{Z}^d\}$$

is a frame (resp. normalized tight frame) for $L^2(\mathbb{R}^d)$. A wavelet frame ψ is called *semi-orthogonal* if $\psi_{m,\ell} \perp \psi_{n,k}$ for all $k, \ell \in \mathbb{Z}^d$ and $m \neq n$. It is called an *orthonormal wavelet* if $\{\psi_{m,\ell}\}$ is an orthonormal basis for $L^2(\mathbb{R}^d)$.

We define the dilation unitary operator D by

$$(Df)(x) = |\det M|^{\frac{1}{2}}\psi(Mx)$$

for $f \in L^2(\mathbb{R}^d)$. Similarly, for any $(s, t) \in \mathbb{R}^{d \times d}$, the translation and modulation unitary operators are defined by

$$T_t f(x) = f(x - t)$$

and

$$E_s f(x) = e^{2\pi i \langle s, x \rangle} f(x)$$

for all $f \in L^2(\mathbb{R}^d)$. Then D, E_s and T_t are unitary operators on $L^2(\mathbb{R}^d)$.

Write $\mathcal{U}_\Lambda = \{E_{\ell_1}T_{\ell_2} : (\ell_1, \ell_2) \in \Lambda\}$ and $\mathcal{U}_{D,T} = \{D^mT_\ell : m \in \mathbb{Z}, \ell \in \mathbb{Z}^d\}$. We will call \mathcal{U}_Λ (resp. $\mathcal{U}_{D,T}$) a *Gabor unitary system* (resp. *wavelet unitary system*). It is easy to check by the definition that the group generated by \mathcal{U}_Λ is contained in $\mathbb{T}\mathcal{U}_\Lambda$, where \mathbb{T} denotes the unit circle.

In quantum and representation theory, a *projective unitary representation* π for a countable discrete (not necessarily abelian) group \mathcal{G} is a mapping $g \rightarrow U_g$ from \mathcal{G} into the set of unitary operators on a Hilbert space H such that $U_gU_h = \mu(g, h)U_{gh}$ for all $g, h \in \mathcal{G}$, where $\mu(g, h)$ belongs to the circle group \mathbb{T} and $(g, h) \rightarrow \mu(g, h)$ is called a *multiplier* of \mathcal{G} ([Va]). It is easy to verify that a Gabor unitary system is the image of a projective unitary representation $\ell \rightarrow U_\ell$ for the group Λ . In general for a countable set of unitary operators \mathcal{U} acting on a separable Hilbert space H that contains the identity operator, we will call \mathcal{U} *group-like* if

$$group(\mathcal{U}) \subset \mathbb{T}\mathcal{U} := \{tU : t \in \mathbb{T}, U \in \mathcal{U}\}$$

and if different U and V in \mathcal{U} are always linearly independent, where $group(\mathcal{U})$ denotes the group generated by \mathcal{U} . We claim that a group-like unitary system \mathcal{U} is always an image of a projective unitary representation π for the group $\mathcal{G} := group(\mathcal{U})$. In fact, for any element $V \in \mathcal{G}$, by the definition of group-like unitary systems there is a unique element $U \in \mathcal{U}$ such that $V = tU$ for some $t \in \mathbb{T}$. Define $\pi(V) = U$. Then V will be a projective unitary representation of \mathcal{G} such that $\pi(\mathcal{G}) = \mathcal{U}$. Typical examples of group-like unitary systems include group unitary systems, Gabor unitary systems and Gabor-type unitary systems (cf. [Ha], [HL]). For Gabor frames we will prove:

Theorem 1.1. *Let Λ be a full rank lattice in $\mathbb{R}^d \times \mathbb{R}^d$, and let g be a Gabor frame generator associated with Λ . Then $S^{-1/2}g$ is the unique best NTF approximation for g , where S is the frame operator for g .*

Theorem 1.1 is in fact a corollary of the following main result of this paper:

Theorem 1.2. *Let \mathcal{U} be a group-like unitary system acting on a Hilbert space \mathcal{H} , and let ψ be a complete frame vector for \mathcal{U} . Then $S^{-1/2}\psi$ is the unique best NTF approximation for ψ , where S is the frame operator for ψ .*

The proofs of Theorem 1.1 and Theorem 1.2 will be given in section 2. We note that Theorem 1.2 is not valid when the group-like unitary system \mathcal{U} is replaced by wavelet unitary systems as the following example shows:

Example 1.1. Suppose that g is an orthonormal M -dilation wavelet. Let $\psi = \frac{1}{4}g$. Then $S^{-1/2}\psi$ is not the best NTF approximation for ψ .

In this example, although $S^{-1/2}\psi$ is not the best NTF approximation among all the normalized tight wavelet frames, it can be checked that it is the best NTF approximation among all the normalized tight ones that are similar to ψ . Therefore this example suggests the following modified NTF approximation problem for wavelet systems:

Question 2. Let ψ be a wavelet frame for $\mathcal{U}_{D,T}$. Does there exist a normalized tight wavelet frame ϕ such that

$$(1) \quad \|\phi - \psi\| = \min\{\|h - \psi\| : h \in \mathcal{T}(\mathcal{U}_{D,T}), h \sim \psi\}?$$

While we are unable to give a full answer to this question, we will prove:

Theorem 1.3. *If ψ is a semi-orthogonal wavelet frame, then there exists a unique normalized tight wavelet frame ϕ such that (1) holds. Moreover, $\phi = S^{-1/2}\psi$.*

In section 3 we will prove Theorem 1.3 and discuss some related problems.

2. GABOR FRAMES

In this section we prove Theorem 1.2, which also implies Theorem 1.1. Since the basic techniques used in this section and the next section involve von Neumann algebra theory, we first introduce some notation.

A *von Neumann algebra* \mathcal{M} is a $*$ -subalgebra of $B(\mathcal{H})$ such that $I \in \mathcal{M}$ and \mathcal{M} is closed in the weak operator (or strong operator) topology. By the double commutant theorem, a $*$ -subalgebra \mathcal{M} of $B(H)$ is a von Neumann algebra if and only if $\mathcal{M} = \mathcal{M}''$, where \mathcal{M}' is the commutant of \mathcal{M} . A von Neumann algebra is said to be *finite* if every isometry in the algebra is unitary. Two projections P and Q in a von Neumann algebra \mathcal{M} are said to be *equivalent* if there is an operator $T \in \mathcal{M}$ such that $TT^* = P$ and $T^*T = Q$. So \mathcal{M} is finite if there is no proper subprojection of I in \mathcal{M} that is equivalent to I . We refer to [KR] for more information about von Neumann algebra theory. For a subset X of H and a subset \mathcal{A} of $B(H)$, we use $[X]$ and $w^*(\mathcal{A})$ to denote the closed subspace generated by X and the von Neumann algebra generated by \mathcal{A} , respectively. In this section we always assume that \mathcal{U} is a group-like unitary system acting on a Hilbert space \mathcal{H} .

Lemma 2.1. *Let ψ be a complete wandering vector for \mathcal{U} . Then a vector $\xi \in \mathcal{H}$ is a complete wandering vector for \mathcal{U} if and only if there is a (unique) unitary operator $A \in \mathcal{U}'$ such that $\xi = A\psi$.*

Proof. Let $C_\psi(\mathcal{U}) := \{T \in B(\mathcal{H}) : TU\psi = UT\psi, U \in \mathcal{U}\}$. Then it is a trivial exercise to check that $C_\psi(\mathcal{U}) = \mathcal{U}'$ since \mathcal{U} is group-like. Thus Lemma 2.1 follows from Proposition 1.3 in [DL]. \square

Even when \mathcal{U} is a wavelet system (in this case $C_\psi(\mathcal{U}) \neq \mathcal{U}'$), the above lemma still gives a parameterization of all the wavelets in terms of the unitary operators in $C_\psi(\mathcal{U})$. This simple observation is essential in the operator-interpolation wavelet theory [DL] developed by Larson and Dai. However, in general, we do not have a complete wandering vector for an arbitrary group-like unitary system \mathcal{U} . To prove Theorem 1.2 (also Theorem 1.3), the von Neumann algebra $w^*(\mathcal{U})$ will play a more important role than its commutant. We will provide a new parameterization for normalized tight frame vectors by the unitary operators in $w^*(\mathcal{U})$. The special case when \mathcal{U} is a group has been discussed in [HL].

Lemma 2.2. *Let \mathcal{M} be a von Neumann algebra on a Hilbert space \mathcal{H} and let $P \in \mathcal{M}'$ be a projection. Suppose that $U \in \mathcal{M}|_{PH}$ is a unitary operator. Then there is a unitary operator $W \in \mathcal{M}$ such that $U = W|_{PH}$.*

Proof. Let $S \in \mathcal{M}|_{PH}$ be a self-adjoint operator B such that $U = e^{iB}$. Then there is an operator $A \in \mathcal{M}$ such that $A|_{PH} = B$. Let $C = \frac{1}{2}(A + A^*)$. Note that $A^*|_{PH} = B^* = B$. It follows that $C|_{PH} = A$. Let $W = e^{iC}$. Then $W \in \mathcal{M}$ is unitary and $W|_{PH} = U$. \square

Let \mathcal{U} be a group-like unitary system on H . By definition, there exists a function $f : \text{group}(\mathcal{U}) \rightarrow \mathbb{T}$ and a mapping $\sigma : \text{group}(\mathcal{U}) \rightarrow \mathcal{U}$ such that $W = f(W)\sigma(W)$ for all $W \in \text{group}(\mathcal{U})$. To see that f and σ are well defined, let $W = \lambda_1 U_1 = \lambda_2 U_2$ with $U_1, U_2 \in \mathcal{U}$ and $\lambda_1, \lambda_2 \in \mathbb{T}$. Then $U_1 = U_2$ and $\lambda_1 = \lambda_2$ since \mathcal{U} is an independent set. Hence both f and σ are well defined. Using this we can define the left (resp. right) regular representation as in the group case.

Let $\ell^2(\mathcal{U})$ be the Hilbert space of square-summable sequences indexed by \mathcal{U} , and let $\{\chi_U\}$ be the standard orthonormal basis where χ_U takes value 1 at U and zero everywhere else. For each fixed $U \in \mathcal{U}$, we define $L_U \in B(\ell^2(\mathcal{U}))$ by the formula

$$L_U \chi_V = f(UV) \chi_{\sigma(UV)}, \quad V \in \mathcal{U}.$$

Then L is a unitary representation of \mathcal{U} on $\ell^2(\mathcal{U})$ such that $L_U L_V = f(UV) L_{\sigma(UV)}$ and $L_U^{-1} = f(U^{-1}) L_{\sigma(U^{-1})}$ for all $U, V \in \mathcal{U}$. In the group case, this is exactly the left regular representation for the group. Therefore we also call L the *left regular representation* for the group-like unitary system \mathcal{U} . The *right regular representation* R_U of \mathcal{U} is defined by

$$R_U \chi_V = f(VU^{-1}) \chi_{\sigma(VU^{-1})}, \quad V \in \mathcal{U}.$$

Let \mathcal{M} denote the von Neumann algebra generated by $\{L_U : U \in \mathcal{U}\}$. As in the group case, the commutant \mathcal{M}' is exactly the von Neumann algebra generated by the right regular representation and both \mathcal{M} and \mathcal{M}' are finite von Neumann algebras ([GH1]). There is a natural conjugate linear isomorphism π from \mathcal{M} onto \mathcal{M}' defined by

$$\pi(A)B\chi_I = BA^*\chi_I, \quad A, B \in \mathcal{M}.$$

In particular, $\pi(A)\chi_I = A^*\chi_I$ for all $A \in \mathcal{M}$. For an orthogonal projection P in \mathcal{M}' , we use $L|_P$ to denote the subrepresentation of L restricted to the range of P .

The following lemma is the main ingredient in proving Theorem 1.2.

Lemma 2.3. *Let η be a complete normalized tight frame vector for a group-like unitary system \mathcal{U} and $\xi \in \mathcal{H}$. Then*

(i) ξ is a complete normalized tight frame vector for \mathcal{U} if and only if there exists a unitary operator $A \in w^*(\mathcal{U})$ such that $A\eta = \xi$.

(ii) ξ is a complete frame vector for \mathcal{U} if and only if there exists an invertible operator $A \in w^*(\mathcal{U})$ such that $A\eta = \xi$.

Proof. We will prove (i), and the proof of (ii) is similar. Let \mathcal{M} be the von Neumann algebra generated by $\{L_U : U \in \mathcal{U}\}$. Define T_η from \mathcal{H} to $\ell^2(\mathcal{U})$ by

$$T_\eta x = \sum_{U \in \mathcal{U}} \langle x, U\eta \rangle \chi_U.$$

Claim. T_η is an isometry such that $T_\eta U = L_U T_\eta$ and $T_\eta \eta = P\chi_I$, where P is the orthogonal projection from $\ell^2(\mathcal{U})$ onto the range space of T_η and $P \in \mathcal{M}'$.

This claim can be checked by using the definitions of group-like unitary systems and normalized tight frame vectors. To save space we leave it to the interested reader or the reader can refer to [GH1] for details.

By the above claim, \mathcal{U} is unitarily equivalent to the unitary system $\{L_U|_P : U \in \mathcal{U}\}$. Therefore, without loss of generality, it suffices to prove this lemma for the case $\tilde{\mathcal{U}} = \{L_U|_P : U \in \mathcal{U}\}$ and $\tilde{\eta} = P\chi_I$, where P is an orthogonal projection in \mathcal{M}' .

First assume that $A \in w^*(\tilde{\mathcal{U}})$ is unitary. Then, by Lemma 2.2, there is a unitary operator $B \in \mathcal{M}$ such that $A = PBP$. So $A\tilde{\eta} = PBP\tilde{\eta} = PBP\chi_I = PB\chi_I$. Now for any $x \in \text{Range}(P)$ we have

$$\begin{aligned} \sum_{\tilde{U} \in \tilde{\mathcal{U}}} |\langle x, \tilde{U}A\tilde{\eta} \rangle|^2 &= \sum_{U \in \mathcal{U}} |\langle x, L_U PB\chi_I \rangle|^2 \\ &= \sum_{U \in \mathcal{U}} |\langle x, PL_U B\chi_I \rangle|^2 \\ &= \sum_{U \in \mathcal{U}} |\langle Px, L_U \pi(B^*)\chi_I \rangle|^2 \\ &= \sum_{U \in \mathcal{U}} |\langle x, \pi(B^*)L_U \chi_I \rangle|^2 \\ &= \sum_{U \in \mathcal{U}} |\langle \pi(B^*)^* x, L_U \chi_I \rangle|^2 \\ &= \|\pi(B^*)^* x\|^2 = \|x\|^2, \end{aligned}$$

where in the fourth equality we use the fact that $\pi(B^*)L_U = L_U\pi(B^*)$, and in the last equality we use the fact that $\pi(B^*)$ is unitary. Therefore $A\tilde{\eta}$ is a complete normalized tight frame vector for $\tilde{\mathcal{U}}$.

Now let $\xi \in \text{Range}(P)$ be a complete normalized tight frame vector for $\tilde{\mathcal{U}}$. We need to find a unitary operator $A \in w^*(\tilde{\mathcal{U}})$ such that $A\tilde{\eta} = \xi$. For this purpose we define a bounded operator B on $\ell^2(\mathcal{U})$ by $B\chi_U = L_U\xi$ ($U \in \mathcal{U}$). Let f and σ be the associated mappings determined by the group-like unitary system \mathcal{U} . Then for any $U, V \in \mathcal{U}$ we have $L_UL_V = f(UV)L_{\sigma(UV)}$. So

$$\begin{aligned} BL_U\chi_V &= Bf(UV)\chi_{\sigma(UV)} = f(UV)B\chi_{\sigma(UV)} \\ &= f(UV)L_{\sigma(UV)}\xi = L_UL_V\xi = L_UB\chi_V. \end{aligned}$$

Hence $BL_U = L_UB$ for all $U \in \mathcal{U}$, which implies that $B \in \mathcal{M}'$.

It is routine to check that $BB^* = P$. Therefore B is a partial isometry in \mathcal{M}' . Let $Q = BB^*$. Then P and Q are equivalent projections in \mathcal{M}' . Since \mathcal{M} is a finite von Neumann algebra, it follows that P^\perp and Q^\perp are also equivalent (cf. [KR]). Therefore there is a partial isometry $C \in \mathcal{M}'$ such that $CC^* = P^\perp$ and $C^*C = Q^\perp$. Let $T = B + C$. Then T is a unitary operator in \mathcal{M}' , and so $P\pi^{-1}(T)P$ is a unitary operator in $w^*(\tilde{\mathcal{U}})$. We also have

$$\begin{aligned} A\tilde{\eta} &= P\pi^{-1}(T)P\chi_I = P\pi^{-1}(T)\chi_I \\ &= P\pi(\pi^{-1}(T))\chi_I = PT\chi_I = P(B + C)\chi_I \\ &= P(\xi + C\chi_I). \end{aligned}$$

Since $P\xi = \xi$ and $\text{Range}(C) = (I - P)\ell^2(\mathcal{U})$, it follows that $A\tilde{\eta} = \xi$, as expected. \square

The following is a simple consequence of the proof of Lemma 2.3 which will be used in the next section.

Corollary 2.4. *Let η be a complete normalized tight frame vector for a group-like unitary system \mathcal{U} . Suppose that $\xi \in \mathcal{H}$ such that $\mathcal{U}\xi$ is a normalized tight frame for the closure $[\mathcal{U}\xi]$ of its linear span. Then there is a partial isometry $A \in W^*(\mathcal{U})$ such that $\xi = A\eta$.*

Proof. As in the proof of Lemma 2.3, we only need to consider the case that $\tilde{\mathcal{U}} = \{L_U|_P : U \in \mathcal{U}\}$, $\tilde{\eta} = P\chi_I$ and $\xi \in \text{Range}(P)$, where P is an orthogonal projection in \mathcal{M}' . Again let B be a bounded operator on $\ell^2(\mathcal{U})$ defined by $B\chi_U = L_U\xi$ ($U \in \mathcal{U}$). Then $B \in \mathcal{M}'$ is a partial isometry. So $\pi^{-1}(B)$ is a partial isometry in \mathcal{M} , which implies that $A := P\pi^{-1}(B)P$ is also a partial isometry in $w^*(\tilde{\mathcal{U}})$. Moreover, a similar calculation as in the proof of Lemma 2.3 shows $A\tilde{\eta} = P\xi$. So $A\tilde{\eta} = \xi$ since $\xi \in \text{Range}(P)$. \square

Proof of Theorem 1.2. Let S be the frame operator for $\mathcal{U}\psi$. We first check that $S \in \mathcal{U}'$. Let f and σ be the associated mappings with \mathcal{U} . Then for any $V \in \mathcal{U}$ and any $x \in \mathcal{H}$ we have

$$\begin{aligned} SVx &= \sum_{U \in \mathcal{U}} \langle Vx, U\psi \rangle U\psi = V \sum_{U \in \mathcal{U}} \langle x, V^*U\psi \rangle V^*U\psi \\ &= V \sum_{U \in \mathcal{U}} \langle x, f(V^*U)\sigma(V^*U)\psi \rangle f(V^*U)\sigma(V^*U)\psi \\ &= V \sum_{U \in \mathcal{U}} \langle x, \sigma(V^*U)\psi \rangle \sigma(V^*U)\psi. \end{aligned}$$

By the definition of group-like unitary systems, it can be checked that $\{\sigma(V^*u) : U \in \mathcal{U}\} = \mathcal{U}$ (cf. [GH1]). So

$$SVx = V \sum_{U \in \mathcal{U}} \langle x, \sigma(V^*U)\psi \rangle \sigma(V^*U)\psi = V \sum_{U \in \mathcal{U}} \langle x, U\psi \rangle U\psi = VSx.$$

Therefore $S \in \mathcal{U}'$. Hence, by the standard spectral decomposition theorem for positive operators, $S^{-1/2}, S^{-1/4} \in \mathcal{U}'$. Hence $\{S^{-1/2}U\psi : U \in \mathcal{U}\} = \{US^{-1/2}\psi : U \in \mathcal{U}\}$, which implies that $\eta := S^{-1/2}\psi$ is a complete normalized tight frame vector for \mathcal{U} .

Now let ξ be any complete normalized tight frame vector for \mathcal{U} . By Lemma 2.3 there is a unitary operator $A \in w^*(\mathcal{U})$ such that $A\eta = \xi$. So

$$\begin{aligned} \|\xi - \psi\|^2 &= \|\xi\|^2 + \|\psi\|^2 - 2\text{Re}\langle \xi, \psi \rangle \\ &= \|A\eta\|^2 + \|\psi\|^2 - 2\text{Re}\langle A\eta, \psi \rangle \\ &= \|\eta\|^2 + \|\psi\|^2 - 2\text{Re}\langle A\eta, \psi \rangle \\ &= \|S^{-1/2}\psi\|^2 + \|\psi\|^2 - 2\text{Re}\langle AS^{-1/2}\psi, \psi \rangle. \end{aligned}$$

Since $AS^{-1/4} = S^{-1/4}A$, it follows from the Cauchy-Schwartz inequality that

$$\begin{aligned} |\text{Re}\langle AS^{-1/2}\psi, \psi \rangle| &= |\text{Re}\langle AS^{-1/4}\psi, S^{-1/4}\psi \rangle| \\ &\leq \|AS^{-1/4}\psi\| \cdot \|S^{-1/4}\psi\| \\ &= \|S^{-1/4}\psi\|^2 = \langle S^{-1/2}\psi, \psi \rangle. \end{aligned}$$

Therefore,

$$\|\xi - \psi\|^2 \geq \|S^{-1/2}\psi\|^2 + \|\psi\|^2 - 2\langle S^{-1/2}\psi, \psi \rangle = \|S^{-1/2}\psi - \psi\|^2.$$

So $S^{-1/2}\psi$ is the best NTF approximation for ψ .

For the uniqueness, assume that ξ is another best NTF approximation for ψ . Then, again by Lemma 2.3, there exists a unitary operator $A \in w^*(\mathcal{U})$ such that $\xi = AS^{-1/2}\psi$. From $\|\xi - \psi\|^2 = \|S^{-1/2}\psi - \psi\|^2$, we have $\operatorname{Re}\langle \xi, \psi \rangle = \langle S^{-1/2}\psi, \psi \rangle$. Note that in the Cauchy-Schwartz inequality, the equality holds if and only if there is a constant α such that $AS^{-1/4}\psi = \alpha S^{-1/4}\psi$. This implies that

$$\xi = AS^{-1/2}\psi = S^{-1/4}AS^{-1/4}\psi = \alpha S^{-1/2}\psi.$$

So

$$\operatorname{Re}\langle \alpha S^{-1/2}\psi, \psi \rangle = \langle S^{-1/2}\psi, \psi \rangle.$$

Since $|\alpha| = 1$, it follows from the above identity that $\alpha = 1$. Thus $\xi = S^{-1/2}\psi$. \square

Proof of Theorem 1.1. Let $\mathcal{H} = L^2(\mathbb{R}^d)$ and $\mathcal{U} = \mathcal{U}_\Lambda$. Then \mathcal{U} is a group-like unitary system. So Theorem 1.1 follows from Theorem 1.2 immediately. \square

Remark. For a full rank lattice Λ in \mathbb{R}^{2d} , we write $\Lambda = M\mathbb{Z}^{2d}$ where M is a non-singular $2d \times 2d$ real matrix. For Gabor systems, one of the fundamental questions is: Under what conditions on Λ do we have a complete frame vector for \mathcal{U}_Λ ? The well-known density theorem (cf. [Dau], [Rie], [RSt]) tells us that one necessary condition is that $|\det M| \leq 1$. In the separable case $\Lambda = \mathcal{L} \times \mathcal{K}$, where \mathcal{L} and \mathcal{K} are full rank lattices in \mathbb{R}^d , D. Han and Y. Wang proved in [HW1] that this condition is also sufficient by solving a problem in the geometry of numbers, motivated by issues in harmonic analysis. It still remains open whether this condition is sufficient in general. For more details about the time-frequency density we refer to [BHW], [CDH], [Dau], [DLL], [FS], [GH1], [GH2], [GH3], [GHL], [Rie], [RSh], [RSt], etc.

Applying Theorem 1.2 to different unitary group systems we can get some interesting examples. For instance,

Example 2.1. Let \mathbb{T} be the unit circle with the normalized Lebesgue measure μ (i.e., $\mu(\mathbb{T}) = 1$). Let E be a measurable subset of \mathbb{T} and let $\mathcal{H} = L^2(E)$. Consider the unitary group $\mathcal{U} = \{M_{e^{imt}} : m \in \mathbb{Z}\}$, where M_f is the multiplication operator by the symbol $f \in L^\infty(E)$. It is easy to prove:

(i) g is a complete normalized tight frame vector for \mathcal{U} if and only if $|g(t)| = 1$, a. e.;

(ii) ψ is a complete frame vector for \mathcal{U} if and only if $0 < a \leq |\psi(t)| \leq b$, a. e., where a, b are constants.

If S is the frame operator associated with a complete frame vector ψ , then $Sf = \sum_m \langle f, e^{imt}\psi \rangle e^{imt}\psi = |\phi|^2 f$. Hence $S = M_{|\psi|^2}$ and so $S^{-1/2} = M_{1/|\psi|}$. Therefore Theorem 1.2 implies that $\frac{\psi}{|\psi|}$ is the unique solution for the following minimization problem:

$$\inf\{\|f - \psi\| : |f(t)| = 1, \text{ a. e.}\}.$$

Moreover,

$$\inf\{\|f - \psi\| : |f(t)| = 1, \text{ a. e.}\} = \|\psi/|\psi| - \psi\|.$$

Example 2.2. Let G be a discrete group and \mathcal{M} be the von Neumann algebra generated by its left regular representation. Define a $\Phi(A) = \langle A\psi, \psi \rangle$ on \mathcal{M} , where $\psi = \chi_e$ with e the identity element in G . Then Φ is a faithful trace on \mathcal{M} , and the Hilbert-Schmidt norm $\|A\|_{HS}$ is defined by $\|A\|_{HS}^2 = \Phi(A^*A)$. Let A be an invertible operator in \mathcal{M} and let $A = V|A|$ be the polar decomposition of A . Then Theorem 1.2 implies the following:

$$\|V - A\|_{HS} = \min\{\|U - A\|_{HS} : U \in \mathcal{U}(\mathcal{M})\},$$

where $\mathcal{U}(\mathcal{M})$ denotes the set of all the unitary operators in \mathcal{M} .

Remark. After this paper was finalized, we were informed that independently and at the same time, Janssen and Strohmer also obtained Theorem 1.1 for the one-dimensional case \mathbb{R} when $\Lambda = \alpha\mathbb{Z} \times \beta\mathbb{Z}$ in [JS] with a completely different approach. However, our results not only apply to general Gabor frames (including Gabor frames for subspaces) but more generally apply to any frames induced by any projective unitary representations and, in particular, to frames induced by translation groups. We thank Professor Strohmer for sending us the manuscript [JS] and also thank professor H. Feichtinger and M. Frank for bringing the work of Janssen and Strohmer to our attention.

3. WAVELET FRAMES

In this section we will mainly prove Theorem 1.3. We start with a few remarks and questions about frame operators and dual wavelet frames. In the theory of wavelet analysis one of the basic problems is to find characterizations for wavelet frames in terms of simple and practical conditions. While there exist characterizations for some special classes of wavelet frames such as orthonormal wavelets and normalized tight wavelet frames (cf. [HW2]), this problem may be completely intractable at this time. A more practical problem [Ca1] asked by P. Casazza is:

Question 3. Characterize the wavelet frame such that it has a dual that is also a wavelet frame.

Let S be the frame operator associated with a wavelet frame ψ . Then by definition $\{S^{-1}D^nT_\ell\psi : n \in \mathbb{Z}, \ell \in \mathbb{Z}^d\}$ is the canonical dual for ψ . Therefore this is a wavelet frame if and only if $S^{-1}D^nT_\ell\psi = D^nT_\ell S^{-1}\psi$ holds for any $n \in \mathbb{Z}, \ell \in \mathbb{Z}^d$, i.e., S^{-1} is in the “local commutant” ([DL]) of $\mathcal{U}_{D,T}$ at ψ :

$$C_\psi(\mathcal{U}_{D,T}) := \{A \in B(L^2(\mathbb{R}^d)) : AD^nT_\ell\psi = D^nT_\ell A\psi, \quad n \in \mathbb{Z}, \ell \in \mathbb{Z}^d\}.$$

In the case that ψ is a Riesz wavelet (i.e., $\{D^nT_\ell\psi\}$ is a Riesz basis), the dual is unique. Hence in this case ψ has a dual which is a wavelet frame if and only if $S^{-1} \in C_\psi(\mathcal{U}_{D,T})$. This condition is far from satisfactory since practically it is very hard to check the condition $S^{-1} \in C_\psi(\mathcal{U}_{D,T})$. Concerning our approximation problem we are interested in the following:

Question 4. Is $S^{-1/2}\psi$ always a wavelet frame? If it is not, under what conditions do we have this property?

Note that $S^{-1/2}\psi$ is a normalized tight wavelet frame if $S^{-1/2} \in C_\psi(\mathcal{U}_{D,T})$. Again this is hard to check. Actually, since $C_\psi(\mathcal{U}_{D,T})$ is not a von Neumann operator algebra (it is a weakly closed linear subspace which contains many interesting von Neumann subalgebras [DL], [La]), we do not even know whether $S^{-1} \in$

$C_\psi(\mathcal{U}_{D,T})$ implies $S^{-1/2} \in C_\psi(\mathcal{U}_{D,T})$. Theorem 1.3 tells us that $S^{-1/2} \in C_\psi(\mathcal{U}_{D,T})$, and so $S^{-1/2}\psi$ is a normalized tight wavelet frame when ψ is semi-orthogonal.

Proof of Theorem 1.3. We first check that $S^{-1/2}\psi$ is a normalized tight wavelet frame. Since $\{S^{-1/2}D^nT_\ell\psi : n \in \mathbb{Z}, \ell \in \mathbb{Z}^d\}$ is a normalized tight frame for $L^2(\mathbb{R}^d)$, it suffices to check that $S^{-1/2}D^nT_\ell\psi = D^nT_\ell S^{-1/2}\psi$ for all $n \in \mathbb{Z}$ and $\ell \in \mathbb{Z}^d$. Write $W_n = [D^nT_\ell\psi]$, the closed subspace generated by $\{D^nT_\ell\psi : \ell \in \mathbb{Z}^d\}$. Then $W_n \perp W_m$ for $m \neq n$, and $\{D^nT_\ell\psi : \ell \in \mathbb{Z}^d\}$ must be a frame for W_n because of the semi-orthogonality of ψ .

Let S_n be the frame operator for $\{D^nT_\ell\psi : \ell \in \mathbb{Z}^d\}$. Then for each $f \in W_m$ we have

$$\begin{aligned} Sf &= \sum_{n \in \mathbb{Z}, \ell \in \mathbb{Z}^d} \langle f, D^nT_\ell\psi \rangle D^nT_\ell\psi \\ &= \sum_{\ell \in \mathbb{Z}^d} \langle f, D^mT_\ell\psi \rangle D^mT_\ell\psi = S_m f \end{aligned}$$

since $f \perp W_n$ for $n \neq m$. Thus $S = \bigoplus_{n \in \mathbb{Z}} S_n$ and hence $S^{-1/2} = \bigoplus_{n \in \mathbb{Z}} S_n^{-1/2}$. Moreover,

$$\begin{aligned} S_n D^n T_\ell \psi &= \sum_{k \in \mathbb{Z}^d} \langle D^n T_\ell \psi, D^n T_k \psi \rangle D^n T_k \psi \\ &= \sum_{k \in \mathbb{Z}^d} \langle T_\ell \psi, T_k \psi \rangle D^n T_k \psi \\ &= D^n \sum_{k \in \mathbb{Z}^d} \langle T_\ell \psi, T_k \psi \rangle T_k \psi \\ &= D^n S_0 T_\ell \psi. \end{aligned}$$

So

$$D^{-n} S_n D^n f = S_0 f$$

for all $f \in W_0$, which implies that $D^{-n} S_n^{-1/2} D^n = S_0^{-1/2}$ on W_0 .

Since $\{T_\ell|_{W_0} : \ell \in \mathbb{Z}^d\}$ is a group, it follows that $S_0^{-1/2} T_\ell \psi = T_\ell S^{-1/2} \psi$. Therefore,

$$\begin{aligned} S^{-1/2} D^n T_\ell \psi &= S_n^{-1/2} D^n T_\ell \psi = D^n S_0^{-1/2} T_\ell \psi \\ &= D^n T_\ell S_0^{-1/2} \psi = D^n T_\ell S^{-1/2} \psi \end{aligned}$$

holds for all $n \in \mathbb{Z}$ and $\ell \in \mathbb{Z}^d$. Thus $S^{-1/2}\psi$ is a normalized tight wavelet frame.

Now we show that $S^{-1/2}\psi$ is a best NTF approximation for ψ among all the normalized tight wavelet frames that are similar to ψ . Suppose that ϕ is any normalized tight wavelet frame that is similar to ψ . Then there is a bounded invertible operator B on $L^2(\mathbb{R}^d)$ such that $BD^nT_\ell\phi = D^nT_\ell\psi$ for all $n \in \mathbb{Z}, \ell \in \mathbb{Z}^d$. Thus

$$S^{-1/2}BD^nT_\ell\phi = S^{-1/2}D^nT_\ell\psi = D^nT_\ell S^{-1/2}\psi.$$

Therefore $S^{-1/2}B$ induces a similarity between the two normalized tight wavelet frames ϕ and $S^{-1/2}\psi$, and so it must be a unitary operator (cf. [HL], Proposition 1.9 (ii)). Let $V = S^{-1/2}B$. Then

$$\begin{aligned}
\langle D^n T_\ell \phi, D^m T_k \phi \rangle &= \langle V D^n T_\ell \phi, V D^m T_k \phi \rangle \\
&= \langle D^n T_\ell S^{-1/2} \psi, D^m T_k S^{-1/2} \psi \rangle \\
&= \langle D^n T_\ell S_0^{-1/2} \psi, D^m T_k S_0^{-1/2} \psi \rangle = 0
\end{aligned}$$

if $m \neq n$ since $W_n \perp W_m$. This implies that ϕ is also semi-orthogonal, and hence $\{T_\ell \phi\}$ is a normalized tight frame for $\overline{\text{span}}\{T_\ell \phi\}$. Let P be the orthogonal projection from $L^2(\mathbb{R}^d)$ onto W_0 . Then $PT_\ell = T_\ell P$ holds for every $\ell \in \mathbb{Z}^d$ since W_0 is an invariant subspace for $\{T_\ell : \ell \in \mathbb{Z}^d\}$. This implies that $\{T_\ell P\phi : \ell \in \mathbb{Z}^d\}$ will be a normalized tight frame for $P\overline{\text{span}}\{T_\ell \phi\}(\subset W_0)$. We also know that $\{T_\ell S^{-1/2} \psi : \ell \in \mathbb{Z}^d\}$ is a normalized tight frame for W_0 , i.e., $S^{-1/2} \psi$ is a complete normalized tight frame vector for the translation group $\{T_\ell : \ell \in \mathbb{Z}^d\}$ restricted to W_0 . Thus, by applying Corollary 2.4 to the unitary group $\{T_\ell|_{W_0} : \ell \in \mathbb{Z}^d\}$, there exists a partial isometry $A \in w^*(T_\ell|_{W_0} : \ell \in \mathbb{Z}^d)$ such that $P\phi = AS^{-1/2} \psi$. In particular, $S_0^{-1/4} A = AS_0^{-1/4}$ since S_0 commutes with $T_\ell|_{W_0}$ for each $\ell \in \mathbb{Z}^d$. Thus

$$\begin{aligned}
|\operatorname{Re}\langle P\phi, \psi \rangle| &= |\operatorname{Re}\langle AS^{-1/2} \psi, \psi \rangle| = |\operatorname{Re}\langle AS^{-1/4} \psi, S^{-1/4} \psi \rangle| \\
&\leq \|AS^{-1/4} \psi\| \|S^{-1/4} \psi\| \leq \|S^{-1/4} \psi\| \|S^{-1/4} \psi\| \\
&= \langle S^{-1/2} \psi, \psi \rangle.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\phi - \psi\|^2 &= \|\phi\|^2 + \|\psi\|^2 - 2\operatorname{Re}\langle \phi, \psi \rangle \\
&= \|V^{-1}S^{-1/2} \psi\|^2 + \|\psi\|^2 - 2\operatorname{Re}\langle \phi, P\psi \rangle \\
&= \|S^{-1/2} \psi\|^2 + \|\psi\|^2 - 2\operatorname{Re}\langle P\phi, \psi \rangle \\
&\geq \|S^{-1/2} \psi\|^2 + \|\psi\|^2 - 2\operatorname{Re}\langle S^{-1/2} \psi, \psi \rangle \\
&= \|S^{-1/2} \psi - \psi\|^2.
\end{aligned}$$

Hence $S^{-1/2} \psi$ is a best NTF approximation for ψ among all the normalized tight wavelet frames that are similar to ψ .

Finally we prove the uniqueness. Assume that ϕ is another NTF approximation for ψ in the problem. In the Cauchy-Schwartz inequality, the equality holds if and only if $AS^{-1/4} \psi = \alpha S^{-1/4} \psi$ for some complex number $\alpha \in \mathbb{C}$. The same argument as in the proof of Theorem 1.2 shows that $\alpha = 1$. Thus $P\phi = S^{-1/2} \psi$, and so $\|P\phi\| = \|S^{-1/2} \psi\| = \|\phi\|$. This implies that $P\phi = \phi$. Thus $\phi = S^{-1/2} \psi$, which implies the uniqueness. \square

Corollary 3.1. *If ψ is a Riesz wavelet such that $S^{-1/2} \in C_\psi(\mathcal{U}_{D,T})$, then $S^{-1/2} \psi$ is the unique best NTF approximation for ψ among those normalized tight wavelet frames that are similar to ψ , where S is the frame operator for ψ .*

Proof. From the proof of Theorem 1.3, it suffices to point out that $\{T_\ell S^{-1/2} \psi\}$ and $\{T_\ell \phi\}$ are normalized tight frames for the closed subspaces they generate, where $\phi \sim \psi$. Since $\phi \sim \psi$ (also $S^{-1/2} \psi \sim \psi$ by assumption), it follows that $\{D^n T_\ell \phi\}$ (resp. $\{D^n T_\ell S^{-1/2} \psi\}$) is also a Riesz basis for $L^2(\mathbb{R}^d)$ since ψ is a Riesz wavelet. It is known that if a normalized tight frame is also a Riesz basis, then it must be an orthonormal basis (cf. [HL], Proposition 1.9 (v)). Therefore $\{D^n T_\ell \phi\}$ (resp. $\{D^n T_\ell S^{-1/2} \psi\}$) is an orthonormal basis for $L^2(\mathbb{R}^d)$. Hence $\{T_\ell S^{-1/2} \psi\}$ and $\{T_\ell \phi\}$ are normalized tight frames for the closed subspaces they generate. \square

Let \mathcal{U} be a unitary system such that $\mathcal{U} = \mathcal{U}_1\mathcal{U}_0$, where \mathcal{U}_1 and \mathcal{U}_0 are two unitary groups such that $\mathcal{U}_1 \cap \mathcal{U}_0 = \{I\}$. Such a \mathcal{U} will be called an *abstract wavelet system*. A complete frame vector ψ for \mathcal{U} is called *semi-orthogonal* if $U\mathcal{U}_0\psi \perp V\mathcal{U}_0\psi$ for different $U, V \in \mathcal{U}_1$. The proof of Theorem 1.3 is clearly valid for the following more general result:

Theorem 3.2. *Let ψ be a semi-orthogonal complete frame vector for an abstract wavelet system \mathcal{U} , and S be the associated frame operator. Then $S^{-1/2}\psi$ is the unique complete normalized tight frame vector such that*

$$\|S^{-1/2}\psi - \psi\| \leq \|\phi - \psi\|$$

holds for every normalized tight frame vector ϕ that is similar to ψ .

Proof of Example 1.1. Let S be the frame operator for ψ . Then a simple argument shows that $S = \frac{1}{16}I$, where I is the identity operator on $L^2(\mathbb{R}^d)$. Thus $S^{-1/2} = 4I$. Therefore,

$$\|S^{-1/2}\psi - \psi\| = 3\|\psi\| = 3/4.$$

Choose ψ_n to be normalized tight wavelet frames such that $\|\psi_n\| \rightarrow 0$ as $n \rightarrow \infty$. The existence of such a sequence can be easily checked. For instance, in the case that $d = 1$ and $A = 2$, we can choose $\hat{\psi}_n = \frac{1}{\sqrt{2\pi}}\chi_{E_n}$ with $E_n = \frac{1}{2^n}([-2\pi, -\pi) \cup [\pi, 2\pi))$. Then ψ_n is a normalized tight wavelet frame (cf. [HL]). Thus,

$$\inf\{\|\phi - \psi\| : \phi \in \mathcal{T}(\mathcal{U}_{D,T})\} \leq \lim_{n \rightarrow \infty} \|\psi_n - \psi\| = \|\psi\| = 1/4.$$

Therefore $S^{-1/2}\psi$ is not the best NTF approximation for ψ . □

ACKNOWLEDGEMENTS

The author would like to thank the referee very much for many very helpful comments and suggestions that helped us improve the presentation of this paper.

REFERENCES

- [AEG1] J. G. Aiken, J. A. Erdos and J. A. Goldstein, Unitary approximation of positive operators, *Illinois J. Math.*, **24** (1980), 61–72. MR **81a**:47026
- [AEG2] J. G. Aiken, J. A. Erdos and J. A. Goldstein, On Löwdin orthogonalization, *Internat. J. Quantum Chem.*, **18** (1980), 1101–1108.
- [Ba] R. Balan, Equivalence relations and distances between Hilbert frames, *Proc. Amer. Math. Soc.*, **127** (1999), 2353–2366. MR **99j**:46025
- [BF] J. Benedetto and M. Fickus, Finite tight frames, *Adv. in Computational Mathematics*, **18** (2003), 357–385.
- [BHW] J. Benedetto, C. Heil and D. F. Walnut, Gabor systems and the Balian-Low theorem, *Gabor analysis and algorithms*, 85–122, Appl. Numer. Harmonic Anal., Birkhäuser Boston, Boston, MA, 1998. MR **98j**:42016
- [BT] J. Benedetto and O. Treiber, Wavelet frames: Multiresolution analysis and extension principles, *Wavelet Transforms and Time-Frequency Signal Analysis*, Ed. L. Debnath, 3–36, Appl. Numer. Harmonic Anal., Birkhäuser Boston, Boston, MA, 2001. MR **2002c**:42048
- [Ca1] P. Casazza, The art of frame theory, *Taiwanese J. Math.*, **4** (2000), 129–201. MR **2001f**:42046
- [Ca2] P. Casazza, Modern tools for Weyl-Heisenberg (Gabor) frame theory, *Adv. Imag. Elect. Phys.*, **115** (2001), 1–127.
- [CK] P. Casazza and J. Kovačević, Uniform tight frames with erasures, preprint.
- [CDH] O. Christensen, B. Deng and C. Heil, Density of Gabor frames, *Appl. Comput. Harmonic Anal.* **7** (1999), 292–304. MR **2000j**:42043

- [DL] X. Dai and D. Larson, Wandering vectors for unitary systems and orthogonal wavelets, *Memoirs Amer. Math. Soc.*, **134** (1998), No. 640. MR **98m**:47067
- [Dau] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Math., SIAM, Philadelphia (1992). MR **93e**:42045
- [DLL] I. Daubechies, H. Landau and Z. Landau, Gabor time-frequency lattices and the Wexler-Raz identity *J. Fourier Anal. Appl.* **1** (1995), 437–478. MR **96i**:42021
- [FS] H. G. Feichtinger and T. Strohmer (eds.), *Gabor Analysis and Algorithms: Theory and Applications, Applied and Numerical Harmonic Analysis*, Birkhäuser, Boston, 1998. MR **98h**:42001
- [FPT] M. Frank, V. I. Paulsen and T. R. Tiballi, Symmetric approximation of frames and bases in Hilbert spaces, *Trans. Amer. Math. Soc.*, **354** (2002), 777–793. MR **2002j**:42042
- [GH1] J. P. Gabardo and D. Han, Frame representations for group-like unitary operator systems, *J. Operator Theory*, to appear.
- [GH2] J. P. Gabardo and D. Han, Subspace Weyl-Heisenberg frames, *J. Fourier Analysis and Appl.*, **7** (2001), 419–433. MR **2002f**:42031
- [GH3] J. P. Gabardo and D. Han, Weyl-Heisenberg dual frames and operator algebras, preprint.
- [GHL] J. P. Gabardo, D. Han and D. Larson, Gabor frames and operator algebras, Wavelet Applications in Signal and Image Processing VIII, *Proc. SPIE.*, **4119** (2000), 337–345.
- [Ga] D. Gabor, Theory of Communication, *J. Inst. Elec. Eng. (London)* **93** (1946), 429–457.
- [GL] J. A. Goldstein and Mel Levy, Linear algebra and quantum chemistry, *Amer. Math. Monthly*, **98** (1991), 710–718. MR **92j**:81366
- [Ha] D. Han, Wandering vectors for irrational rotation unitary systems, *Trans. Amer. Math. Soc.*, **350** (1998), 309–320. MR **98k**:47089
- [HL] D. Han and D. Larson, Frames, bases and group representations, *Memoirs Amer. Math. Soc.*, **147** (2000). MR **2001a**:47013
- [HW1] D. Han and Y. Wang, Lattice tiling and the Weyl-Heisenberg frames, *Geometric and Functional Analysis*, **11** (2001), 742–758.
- [HW2] E. Hernández and G. Weiss, *A First Course on Wavelets*, CRC Press, Boca Raton, FL, 1996. MR **97i**:42015
- [JS] A. Janssen and T. Strohmer, Characterization and computation of canonical tight windows for Gabor frames, *J. of Fourier Analysis and Appl.*, **8** (2002), 1–28. MR **2002m**:42040
- [KR] R. Kadison and J. Ringrose, *Fundamentals of the Theory of Operator Algebras*, Vols. I and II, Academic Press, Inc. 1983 and 1986. MR **85j**:46099; MR **88d**:46106
- [La] D. Larson, von Neumann algebras and wavelets, *Proc. NATO Adv. Studies*, 495 (1997), 267–312. MR **98g**:46091
- [Lo] P. -O. Löwdin, On the nonorthogonality problem, *Adv. Quantum Chem.*, **5**(1970), 185–199.
- [Rie] M. A. Rieffel, von Neumann algebras associated with pairs of lattices in Lie groups, *Math. Ann.* **257** (1981), 403–418. MR **84f**:22010
- [RSh] A. Ron and Z. Shen, Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$, *Duke Mathematical Journal* **89** (1997), 237–282. MR **98i**:42013
- [RSt] J. Ramanathan and T. Steger, Incompleteness of sparse coherent states, *Appl. Comput. Harmonic Anal.* **2** (1995), 148–153. MR **96b**:81049
- [Va] V. S. Varadarajan, *Geometry of Quantum Theory*, Second Edition, Springer-Verlag, New York-Berlin, 1985. MR **87a**:81009

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CENTRAL FLORIDA, ORLANDO, FLORIDA 32816
 E-mail address: dhan@pegasus.cc.ucf.edu

MEAN CURVATURE FLOW, ORBITS, MOMENT MAPS

TOMMASO PACINI

ABSTRACT. Given a compact Riemannian manifold together with a group of isometries, we discuss MCF of the orbits and some applications: e.g., finding minimal orbits. We then specialize to Lagrangian orbits in Kaehler manifolds. In particular, in the Kaehler-Einstein case we find a relation between MCF and moment maps which, for example, proves that the minimal Lagrangian orbits are isolated.

1. INTRODUCTION

If a given submanifold Σ in a Riemannian manifold (M, g) is not minimal, “mean curvature flow” (MCF) provides a canonical way to deform it.

Ideally, the flow should exist until either a singularity develops, preventing further flow, or the submanifold becomes minimal. In this sense, MCF should be a useful tool in the search for minimal submanifolds.

A third, but exceptional, possibility that might occur is exemplified by the “translating solitons” in \mathbb{R}^n : submanifolds which, under MCF, simply flow by translation and thus never converge to a limiting object.

However, MCF is a difficult topic with many open questions. In particular, there is no general theory which can explain what will happen to all $\Sigma \subseteq (M, g)$ under MCF, or classify which singularities can arise. One is thus forced to study each case individually, or at best to look for classes of submanifolds which, under MCF, behave in the same way. Most often this leads to restrictions on the dimension (e.g., curves) or codimension (e.g., hypersurfaces) of Σ , or on properties of the immersion (e.g., convex).

In general, the presence of symmetries in a problem reduces the number of variables, hopefully making things easier. Regarding MCF, the “best” case is when Σ is the orbit of a group of isometries of (M, g) . It is simple to show that, in this case, all Σ_t obtained by MCF are also orbits, and MCF basically reduces to solving an ODE on the (finite-dimensional) space of orbits.

Group actions have been extensively studied. In particular, orbits of a (compact, connected) Lie group G acting on a (compact, connected) manifold M can be classified into three categories: “principal”, “exceptional” and “singular”. This yields a simple and pretty picture of the geometry of the orbit space M/G .

The first goal of this paper is to fit MCF into this framework, analyzing “what happens” to a principal (or exceptional, or singular) orbit under MCF.

The final picture, presented in Theorem 2, constitutes, for several reasons, a good “example zero” of MCF. It is simple; it generalizes the standard example of

Received by the editors September 4, 2002 and, in revised form, January 29, 2003.
 2000 *Mathematics Subject Classification.* Primary 53C42, 53C44; Secondary 53D20.

the “shrinking sphere” in \mathbb{R}^n ; it is (co)dimension-independent; and especially, in the orbit setting, “everything we might want to be true for MCF, is true”.

We then restrict our attention to Lagrangian orbits. Using moment maps to “get a grasp on them”, we prove that the set $\mathcal{L}(M; G)$ of points belonging to Lagrangian G -orbits constitutes a smooth submanifold in any compact Kaehler ambient space. When (M, g) is a Kaehler-Einstein (KE) manifold, it was already known that the Lagrangian condition is preserved under MCF; we give an independent proof of this in the orbit setting, and are then free to apply Theorem 2 to study how Lagrangian orbits evolve under MCF. As a simple corollary, we find that “backwards MCF” always leads to a minimal Lagrangian orbit.

Futaki proved that compact positive KE manifolds come with a “canonical” moment map μ . In Proposition 5 we show that μ is intimately related to the mean curvature of Lagrangian orbits. As a corollary, we find that minimal Lagrangian orbits (wrt fixed G) are isolated.

As already noted, MCF of Lagrangian submanifolds is not a new subject; there is also some overlap, in the case of torus actions, with [G]. However, given the number of known KE manifolds with large isometry groups, there seems to be no a priori reason to limit oneself to tori. Our attempt is to develop a “complete” picture of the general G case, relying only on the basic tools provided by the general theory of G -actions, moment maps and transformation groups. In this sense, we are not aware of any serious overlap with existent literature.

2. SMOOTH GROUP ACTIONS ON MANIFOLDS

This section is mostly a review of standard facts regarding manifolds with a group action. We refer to [A] for further details.

We adopt the following conventions.

- M is a compact, connected, smooth manifold. $\text{Diff}(M)$ will denote its group of diffeomorphisms.
- G is a compact, connected, Lie group acting on M ; i.e., we are given a homomorphism $i : G \longrightarrow \text{Diff}(M)$. The action of $g \in G$ on $p \in M$ will be denoted $g \cdot p$.

The action is “effective” if i is injective. Notice that, since $\text{Ker}(i)$ is normal in G , we may “reduce” any G -action to an effective $G/\text{Ker}(i)$ -action.

Whenever a group H is not connected, H^0 will denote the connected component containing the identity element.

The action of G on M induces an action of G on TM . If $X \in T_p M$, it is defined as follows:

$$g \cdot X := g_*[p](X) \in T_{gp}M$$

where g_* denotes the differential of the map $g = i(g) : M \longrightarrow M$.

Letting \mathfrak{g} denote the Lie algebra of G , any $X \in \mathfrak{g}$ induces a “fundamental vector field” \tilde{X} on M , defined as follows:

$$\tilde{X}(p) := \frac{d}{dt}[\exp(tX) \cdot p]_{t=0}$$

where $\exp(tX)$ denotes the 1-dimensional subgroup of G associated to X .

For all $p \in M$, we define:

$$\begin{aligned} G \cdot p &:= \{g \cdot p : g \in G\} \subseteq M && \text{“orbit of } p \text{ (wrt } G\text{)”}, \\ G_p &:= \{g \in G : g \cdot p = p\} \leq G && \text{“stabilizer of } p \text{ (wrt } G\text{)”}. \end{aligned}$$

Notice that G_p is a closed subgroup of G and that $G \cdot p \simeq G/G_p$. A generic orbit will often be denoted \mathcal{O} . Thanks to the compactness hypotheses, every orbit \mathcal{O} is a smooth embedded submanifold of M .

Notice also that $G_{hp} = h \cdot G_p \cdot h^{-1}$; i.e., G_{hp} is conjugate to G_p . Thus, to each $p \in M$ we may associate a conjugacy class of subgroups of G :

$$\mathcal{O} \mapsto (G_p) := \text{conjugacy class of the stabilizer of any } p \in \mathcal{O}.$$

This class is called the "type" of \mathcal{O} .

Let \mathcal{O} be any orbit and $p \in \mathcal{O}$. Let $V := T_p M / T_p \mathcal{O}$. The action of G on TM restricts to an action of G_p on $T_p M$, and $T_p \mathcal{O} \leq T_p M$ is a G_p -invariant subspace. Thus there is a natural action of G_p on V .

This induces an action of G_p on $G \times V$, as follows:

$$h \cdot (g, v) := (g h^{-1}, h \cdot v).$$

Let $G \times_{G_p} V := (G \times V) / G_p$ denote the quotient space. Then $G \times_{G_p} V$ is a vector bundle (with fiber V) over $G/G_p \simeq \mathcal{O}$ and there is an action of G on $G \times_{G_p} V$ as follows:

$$g_1 \cdot [g_2, v] := [g_1 g_2, v].$$

The following result shows that $G \times_{G_p} V$ contains complete information on the local geometry of the group action near \mathcal{O} .

Theorem 1. *Let G, M be as above.*

Then there exist a G -invariant neighborhood U of \mathcal{O} in M and a G -invariant neighborhood W of the zero section of $G \times_{G_p} V$ such that U is G -equivariantly diffeomorphic to W .

Corollary 1. *Let M, G be as above.*

- (1) *For each fixed type, the union of all orbits of that type forms a (possibly disconnected) submanifold of M .*
- (2) *There is only a finite number of orbit types.*
- (3) *There is an orbit type (P) whose orbits occupy an open, dense, connected subset of M .*

The types of the G -action can be partially ordered by the following relation:

$$\alpha \leq \beta \Leftrightarrow \exists H, K \leq G : \alpha = (H), \beta = (K), H \leq K.$$

If a given orbit \mathcal{O} has type (K) , any nearby orbit $\mathcal{O}' \subset G \times_K V$ can be written $\mathcal{O}' = G \cdot [1, v]$; it is simple to show that the stabilizer of $[1, v]$ is the stabilizer K_v of $v \in V$ wrt the K -action; so it is a subgroup of K . In other words, $\text{type}(\mathcal{O}') \leq \text{type}(\mathcal{O})$.

In particular, the type (P) defined by Corollary 1 must be an absolute minimum:

$$(P) \leq (K), \text{ for all types } (K).$$

It is also clear that $\dim \mathcal{O}' \geq \dim \mathcal{O}$ (the dimension of orbits is a lower-semicontinuous function on M) and that orbits of type (P) have maximum dimension among all orbits.

The final picture is thus as follows.

Given M, G as above, there are three categories of orbits:

- (1) "Principal orbits", corresponding to the minimal type (P) .

They occupy an open, dense, connected subset of M .

- (2) “Exceptional orbits”, corresponding to those types (K) : K/P is finite. Via the projection $G \times_K V \longrightarrow G/K$, any nearby principal orbit is a finite covering of the exceptional orbit G/K .

In particular, exceptional orbits and principal orbits have the same dimension.

- (3) “Singular orbits”, corresponding to those types (K) : $\dim K > \dim P$.

Their dimension is strictly smaller than that of principal orbits.

Example 1. S^1 acts isometrically on $S^2 := \{x^2 + y^2 + z^2 = 1\} \subseteq \mathbb{R}^3$ by rotations along the z -axis.

The orbits are the sets $S^2 \cap \pi_c$, where $\pi_c := \{z \equiv c\}$. The singular orbits, of type (S^1) , are the poles; all other orbits are principal, of type (1).

Example 2. On S^2 there is also an isometric \mathbb{Z}_2 -action that identifies antipodal points. Since the two actions commute, the S^1 -action passes to the quotient $\mathbb{RP}^2 \simeq S^2/\mathbb{Z}_2$.

There is one singular orbit, represented by $S^2 \cap \pi_1$; one exceptional orbit, represented by $S^2 \cap \pi_0$ (the “equator”); all other orbits, represented by $S^2 \cap \pi_c$, $0 < c < 1$, are principle.

Together, principal and exceptional orbits constitute the set of “regular orbits”. Any regular orbit $\mathcal{O} = G \cdot q$, $q \in M^{reg}$ is the image of an immersion

$$\phi : G/P \hookrightarrow M, \quad [g] \mapsto [g] \cdot q.$$

Notice that, if \mathcal{O} is principle, then $\mathcal{O} \simeq G/P$, i.e., ϕ is an embedding. If \mathcal{O} is exceptional, of type (K) , then $\mathcal{O} \simeq G/K$ and ϕ is a covering map of G/P over G/K .

We may set $M^{pr} := \{p \in M : G \cdot p \text{ is a principal orbit}\} \subseteq M$ and, analogously, define M^{ex} , M^{sing} , $M^{reg} = M^{pr} \cup M^{ex}$.

Each of these subsets, generically denoted M^* , is a smooth submanifold inside M and M^*/G also has a smooth structure.

Thus the set M/G , which is compact and Hausdorff with respect to the quotient topology, has the structure of a “stratified smooth manifold”, the smooth strata being the connected components of M^*/G . Once again, M^{pr}/G occupies an open, dense, connected subset of M/G .

An interesting application of all of the above is the following, simple, fact.

Corollary 2. Assume G acts on M , with principal type (P) .

- (1) If P is normal, then $P \leq G_p$, $\forall p \in M$.

Thus, if the action is effective, $P = \{1\}$.

- (2) If P^0 is normal, then $P^0 \leq G_p$, $\forall p \in M$.

Thus, if the action is effective, P is finite.

In particular, assume a torus T acts effectively on M . Then $(P) = \{1\}$.

Proof. If P is normal, $P \leq G_p$, $\forall p \in M^{pr}$. Since M^{pr} is dense in M , it is easy to prove that $P \leq G_p$, $\forall p \in M$.

The proof of (2) is similar. □

Our last goal, in this section, is to “understand” convergence of orbits.

Assume given a curve of principal orbits \mathcal{O}_t (corresponding to immersions $\phi_t : G/P \hookrightarrow M$) which, in the topology of M/G , converges to some limiting orbit \mathcal{O} . We must distinguish three cases.

- (1) Assume $\mathcal{O} \simeq \phi : G/P \hookrightarrow M$ is principal, i.e., has minimal type (P) .

Since M^{pr} is open in M , each orbit near \mathcal{O} must also have type (P) ; thus, wrt the local linearization $M = G \times_P V$ based at \mathcal{O} , P acts trivially on V and $G \times_P V = G/P \times V$ is the trivial bundle. In particular, this shows that $\mathcal{O}_t \rightarrow \mathcal{O}$ smoothly in M , i.e., $\phi_t \rightarrow \phi$.

- (2) Assume $\mathcal{O} \simeq \phi$ is exceptional. Then, near \mathcal{O} , there are either exceptional or principal orbits and they are coverings of \mathcal{O} . It is still true that $\phi_t \rightarrow \phi$ smoothly, but the limit is not injective.
- (3) Assume \mathcal{O} is singular. Let K be the stabilizer of $p \in \mathcal{O}$.

Locally, $M = G \times_K V$, $\mathcal{O}_t = G \cdot [1, v_t]$ (for some $v_t \in V$) and $K_{v_t} \leq K$ is the stabilizer of $[1, v_t]$. Since $(K_{v_t}) \equiv (P)$, all K_{v_t} have constant dimension q . The corresponding Lie algebras \mathfrak{k}_{v_t} are thus points in the Grassmannian $Gr(q, \mathfrak{k})$ of q -planes in $\mathfrak{k} := Lie(K)$. By compactness of $Gr(q, \mathfrak{k})$, we may conclude the following: any sequence $\mathcal{O}_n \subseteq \mathcal{O}_t$, $\mathcal{O}_n \rightarrow \mathcal{O}$, contains a subsequence \mathcal{O}_{n_k} such that $\mathfrak{k}_{n_k} \rightarrow \mathfrak{k}_0$, for some $\mathfrak{k}_0 \in Gr(q, \mathfrak{k})$.

Let $\{X_1, \dots, X_r\}$ span a complement of \mathfrak{k}_0 in \mathfrak{k} , and let $\{Y_1, \dots, Y_s\}$ span a complement of \mathfrak{k} in \mathfrak{g} . Then $T_p \mathcal{O}$ is generated by the fundamental vector fields \tilde{Y}_i , and $T_{[1, v_{n_k}]} \mathcal{O}_{n_k}$ is generated by \tilde{X}_i and \tilde{Y}_j . Since \tilde{X}_i are smooth on M and $\tilde{X}_i|_{\mathcal{O}} \equiv 0$, we see that $\|\tilde{X}_i|_{\mathcal{O}_{n_k}}\| \rightarrow 0$ (wrt any invariant metric on M).

In other words, convergence to a singular orbit is described, up to subsequences, by the vanishing of certain fundamental vector fields; which fields vanish depends on the particular subsequence.

3. MCF OF ORBITS

Let us now fix a compact, connected, Riemannian manifold (M, g) and a compact, connected, Lie group of isometries, $G \leq Isom_g(M)$. (P) will denote the minimal type of the G -action, and \mathfrak{p} the corresponding Lie algebra.

Recall that, to any immersion $\phi : \Sigma \hookrightarrow M$, we may associate a volume

$$vol(\phi) := \int_{\Sigma} vol_{\phi^*g}.$$

Since any regular orbit \mathcal{O} corresponds to an immersion $\phi : G/P \hookrightarrow M$, we get a function

$$vol : (M/G)^{reg} \longrightarrow \mathbb{R}, \quad \mathcal{O} \mapsto vol(\phi).$$

The quotient map $\pi : M \longrightarrow M/G$ yields a pull-back map $\pi^*vol : M^{reg} \longrightarrow \mathbb{R}$. We will often write $vol(\mathcal{O})$ instead of $vol(\phi)$ and vol instead of π^*vol .

Proposition 1. *The volume function has the following properties:*

- (1) $vol : M^{reg} \longrightarrow \mathbb{R}$ is smooth.
- (2) It has a continuous extension to zero on M^{sing} .

This defines a continuous function $vol : M \longrightarrow \mathbb{R}$.

- (3) The function $vol^2 : M \longrightarrow \mathbb{R}$ is smooth.

Proof. For any regular orbit $\mathcal{O} = \phi : G/P \hookrightarrow M$, ϕ^*g defines a G -invariant metric on G/P . Let Z_1, \dots, Z_n be any basis of $T_{[1]}G/P$, induced by the projection onto $\mathfrak{g}/\mathfrak{p}$ of elements $Z_i \in \mathfrak{g} : Z_i \notin \mathfrak{p}$. Let $\mu := Z_1^* \wedge \dots \wedge Z_n^*$ be the induced left-invariant volume form on G/P .

Since both volume forms are G -invariant, $\text{vol}_g = c \cdot \mu$, for some $c = c(\mathcal{O})$. Clearly, $c = \sqrt{\det g_{ij}}$, where $g_{ij} := \phi^* g[1](Z_i, Z_j)$. Thus

$$\text{vol}(\mathcal{O}) = \int_{G/P} \text{vol}_{\phi^* g} = \int_{G/P} \sqrt{\det g_{ij}} \cdot \mu = \sqrt{\det g_{ij}} \cdot \int_{G/P} \mu.$$

Now let \mathcal{O}_t be a curve of regular orbits, $\mathcal{O}_t = \phi_t : G/P \hookrightarrow M$. Assume that $\mathcal{O}_t \rightarrow \mathcal{O}$. If \mathcal{O} is also regular, $\mathcal{O} = \phi : G/P \hookrightarrow M$, then $p_t := \phi_t[1] \rightarrow p := \phi[1]$. Choose $Z_i \in \mathfrak{g}$ such that the induced fundamental vector fields \tilde{Z}_i span $T_p \mathcal{O}$. Then \tilde{Z}_i also span $T_{p_t} \mathcal{O}_t$. Setting $g_{ij}^t := g[p_t](\tilde{Z}_i, \tilde{Z}_j)$ and $g_{ij} := g[p](\tilde{Z}_i, \tilde{Z}_j)$, we find

$$\text{vol}(\mathcal{O}_t) = \sqrt{\det g_{ij}^t} \cdot \text{constant} \longrightarrow \sqrt{\det g_{ij}} \cdot \text{constant} \neq 0.$$

This shows that vol is smooth on M^{reg} . If \mathcal{O} is singular, we saw in section 2 that, for any sequence $\mathcal{O}_n \subseteq \mathcal{O}_t : \mathcal{O}_n \rightarrow \mathcal{O}$, we may choose Z_i so that, for some subsequence, certain \tilde{Z}_i vanish. This shows that $\sqrt{\det g_{ij}^t} \rightarrow 0$; so vol extends continuously to zero on M^{sing} .

Since $\text{vol}(\mathcal{O}_t)^2 = \det g_{ij}^t \cdot \text{constant}$, vol^2 is smooth on M . □

Corollary 3 (cf. [H]). *Let G be any compact, connected Lie group acting by isometries on a compact, connected Riemannian manifold (M, g) .*

Then there exists a regular minimal orbit of the G -action.

Proof. Since M is compact, the continuous function $\text{vol} : M \longrightarrow \mathbb{R}$ has a maximum, which necessarily corresponds to a minimal (immersed) orbit. □

Example 2 of section 2 shows that the minimal orbit might be exceptional.

Let us now recall the notion of “mean curvature flow”.

Fix manifolds Σ and (M, g) , and an immersion $\phi : \Sigma \hookrightarrow M$.

A smooth 1-parameter family of immersions $\phi_t : \Sigma \hookrightarrow M$ is called a “solution to the MCF of (Σ, ϕ) ” if it satisfies the following equation:

$$\text{(MCF)} \quad \begin{cases} \frac{d}{dt} \phi_t &= H(\phi_t), \\ \phi_0 &= \phi, \end{cases}$$

where $H(\phi_t)$ denotes the “mean curvature vector field” of ϕ_t , defined as the trace of the second fundamental form of the immersion. It is well known that H is, up to sign, the “ L^2 -gradient” of the volume functional on immersions:

$$\frac{d}{dt} \text{vol}(\phi_t)|_{t=0} = - \int_{\Sigma} (H, \frac{d}{dt} \phi_{t|_{t=0}}).$$

Locally, (MCF) can be written as a II-order quasi-linear parabolic system of equations. In particular, solutions always exist for some short time interval $t \in [0, \epsilon)$ and are unique.

We want to focus on solving (MCF) under the assumption that (Σ, ϕ) is an orbit of a group of isometries.

Consider the map $H : p \mapsto H_p$ that associates to each $p \in M$ the mean curvature H_p of the orbit $G \cdot p$. This defines a vector field on M .

The following lemma examines its continuity/smoothness.

Lemma 1. *Let H be defined as above.*

- (1) *H is smooth along each submanifold given by orbits of the same type.*
It is also smooth on M^{reg} .

(2) H is G -invariant.

Proof. The smoothness of H along orbits of the same type is clear. Smoothness on M^{reg} comes from the convergence properties of regular orbits: basically, H is a local object and does not notice the difference between principal and exceptional orbits.

(2) is a consequence of the fact that all ingredients in the definition of H are G -invariant. \square

In particular, H descends to a vector field on M/G and is smooth along each stratum. If $\mathcal{O} = \phi : G/K \hookrightarrow M$ and $p := \phi([1])$, we can consider the following ODE on M :

$$(MCF') \quad \begin{cases} \dot{p}(t) &= H[p(t)], \\ p(0) &= p. \end{cases}$$

Notice that, given a solution $p(t)$ of (MCF'), the G -equivariant map

$$\phi : G/K \times [0, \epsilon) \longrightarrow M, \quad \phi([g], t) := g \cdot p(t)$$

solves (MCF) with the initial condition $(\Sigma, \phi) = \mathcal{O}$. By uniqueness of solutions of (MCF), this shows that MCF of an orbit gives a curve of orbits.

In other words, if (Σ, ϕ) is an orbit, (MCF) is equivalent to the ODE on M/G (or on M) determined by integrating H .

The reduction of the problem from a PDE to an ODE simplifies things enormously. For example, MCF of orbits has the following properties:

- There exists a (unique) solution \mathcal{O}_t defined on a maximal time interval (α, β) : this comes from standard ODE theory.
- (MCF) may be inverted; i.e., $t \mapsto \mathcal{Q}(t) := \mathcal{O}(-t)$ solves the equation for "backward MCF":

$$\frac{d}{dt} \mathcal{Q}_t = -H_{\mathcal{Q}_t}, \quad \mathcal{Q}_0 = \mathcal{O}.$$

This is true for any ODE of the type $\dot{x} = f(x(t))$, but is very atypical for parabolic problems.

Another interesting feature of (MCF) on orbits is that it preserves types:

Proposition 2. *For each orbit \mathcal{O} , $H_{\mathcal{O}}$ is tangent to the submanifold determined by the type of \mathcal{O} .*

In particular, if \mathcal{O}_t is the solution of (MCF) with initial condition $\mathcal{O}_0 = \mathcal{O}$, then $\text{type}(\mathcal{O}_t) \equiv \text{type}(\mathcal{O})$.

Proof. Let $p \in M$ and let $G \cdot p$ have stabilizer K . Locally near $G \cdot p$, $M = G \times_K V$, where $V = T_p(G \cdot p)^\perp$ and K acts isometrically on V .

This determines a decomposition of V into K -irreducible subspaces: $V = \bigoplus V^i$. Since H is G -invariant, it is also K -invariant; so $H \in V^0 := \{v \in V : k \cdot v = v, \forall k \in K\}$.

Notice that $G \times_K V^0$ corresponds to the orbits near $G \cdot p$ of type (K) . Thus H_p is tangent to the set of such orbits.

Since this is true for each $p \in M$, (MCF) preserves types. \square

Corollary 4 (cf. [HL]). *Let (M, g) be as above.*

If an orbit is isolated wrt all other orbits of the same type, then it is minimal.

We now want to show that, on M^{reg} , (MCF) is actually a gradient flow; i.e., for some $f \in C^\infty(M^{reg})$, $H = \nabla f$.

Let p_t be a curve in M^{reg} and $\mathcal{O}_t := G \cdot p_t$. We will let X denote both the vector $\frac{d}{dt}p_t|_{t=0}$ at p_0 and the G -invariant vector field $\frac{d}{dt}\mathcal{O}_t|_{t=0}$ along \mathcal{O} .

Since H, X and the metric on M are G -invariant, (H, X) also is. Thus:

$$\frac{d}{dt} \text{vol}(p_t)|_{t=0} = \frac{d}{dt} \text{vol}(\mathcal{O}_t)|_{t=0} = - \int_{G/P} (H, X) \text{vol}_{\mathcal{O}} = -\text{vol}(\mathcal{O}) \cdot (H, X)_{p_0}.$$

This proves that

$$(H, X)_{p_0} = - \frac{\frac{d}{dt} \text{vol}(p_t)|_{t=0}}{\text{vol}(p_0)} = - \frac{d}{dt} \log \text{vol}(p_t)|_{t=0}.$$

In other words, $H = -\nabla \log(\text{vol})$ on M^{reg} .

We now have all the information we need to understand how MCF fits into the framework set up in section 2.

Let $\mathcal{O} = \phi : G/P \hookrightarrow M$ be a fixed principal orbit and let $\mathcal{O}_t = \phi_t : G/P \hookrightarrow M$ be the maximal curve obtained by MCF, with initial condition $\mathcal{O}(0) = \mathcal{O}$ and $t \in (\alpha, \beta)$. Let $p_t = \phi_t([1])$, so that $\frac{d}{dt}p_t = H[p_t]$.

Since M is compact, there is a sequence $\{p_n\} \subseteq \{p_t\}$ and $\tilde{p} \in M$ such that $p_n \rightarrow \tilde{p}$. Thus $\mathcal{O}_n := G \cdot p_n \rightarrow \tilde{\mathcal{O}} := G \cdot \tilde{p}$.

In general, however, different sequences may have different limits; so we cannot hope that $\mathcal{O}_t \rightarrow \tilde{\mathcal{O}}$. The following example of this was suggested to the author by T. Ilmanen.

Example 3. Consider an embedding $s : \mathbb{R} \hookrightarrow \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > 1\}$ that tends towards $S^1 = \{x^2 + y^2 = 1\} \subseteq \mathbb{R}^2$ as $t \rightarrow \infty$, spiralling around it.

Let S denote its image and $f : S \rightarrow \mathbb{R}^1$ be a positive, decreasing function on S such that $f(t) \searrow c > 0$ as $t \rightarrow \infty$.

S may be “fattened” by a tubular neighborhood U (of decreasing width, as $t \rightarrow \infty$). At each point $s \in S$, f may be extended, with constant value $f(s)$, in the normal directions. This gives an extension of $f : U \rightarrow \mathbb{R}$ such that $\nabla f|_S$ is tangent to S . A partition of unity argument now allows us to extend f to a smooth function $\tilde{f} : \mathbb{R}^2 \rightarrow \mathbb{R}$. Clearly, $\nabla \tilde{f}|_S = \nabla f|_S$ and $\tilde{f}|_{S^1} \equiv c$.

Since we are only interested in what happens near S^1 , we may perturb \tilde{f} so that it extends to some compact (Σ^2, g) containing a neighborhood of S^1 . We have thus built a smooth gradient vector field on a compact manifold whose flow does not converge to a unique point.

Now let $M := \Sigma \times S^1$, with the obvious S^1 -action. For each $p \in \Sigma$, let $\{p\} \times S^1$ have an S^1 -invariant metric h_p such that $\log \text{vol}(\{p\} \times S^1) = f(p)$.

For each $(p, q) \in M$, let $T_{p,q}M$ have the product metric $g[p] \times h_p[q]$.

Since $H_{\mathcal{O}} = -\nabla \log \text{vol}(\mathcal{O}) = -\nabla f$, MCF of any orbit $\{s\} \times S^1$ with $s = s(t_0) \in S$ yields the curve of orbits $\{s(t)\} \times S^1$, $t \in [t_0, \infty)$. By construction, these orbits have no limit as $t \rightarrow \infty$.

We are thus interested in conditions ensuring the existence of $\lim_{t \rightarrow \beta} \mathcal{O}_t$.

The following lemma shows that, in the analytic context, things work nicely.

Lemma 2. Consider the ODE

$$\dot{p} = -\nabla f, \quad p(0) = p_0$$

with maximal solution $p(t)$, $t \in (\alpha, \beta)$.

Assume that, for some subsequence $t_n \rightarrow \beta$, $p(t_n) \rightarrow y$ and that f is analytic in a neighborhood of y . Then $p(t) \rightarrow y$.

Proof. Let $s(t) := \int_0^t \|\nabla f(p_\tau)\| d\tau$. Since $\|\nabla f(p_\tau)\| > 0$, s is a diffeomorphism between (α, β) and some (α', β') .

Since $f(p_t)$ is monotone, it also gives a diffeomorphism between (α, β) and some (a, b) , where $b = f(y)$. In particular, s can be written as a function of $f \in (a, b)$; i.e., $s = s(f)$.

Since $\frac{df}{dt} = \nabla f \cdot \frac{dp}{dt} = -|\nabla f|^2$, we find that $\frac{df}{ds} = -|\nabla f|$ and $\frac{ds}{df} = -\frac{1}{|\nabla f|}$.

It is simple to prove that $p(t) \rightarrow y \Leftrightarrow \beta' < \infty$.

We may assume that $f(y) = 0$. When f is analytic near y , the ‘‘Lojasiewicz inequality’’ asserts that there exists a neighborhood U of y and $\delta > 1$, $c > 0$ such that, on U , $|f| \leq c|\nabla f|^\delta$. Thus $\frac{1}{|\nabla f|} \leq c \cdot |f|^{-\frac{1}{\delta}}$, and so

$$\beta' = \int_{f(p_0)}^{f(y)} \dot{s}(f) df = \int_b^{f(p_0)} \frac{1}{|\nabla f|} df = \int_0^{f(p_0)} \frac{1}{|\nabla f|} df < \infty.$$

□

If (M, g) is analytic and $\text{vol}(\mathcal{O}_t) \geq c > 0$, we may apply this lemma to $f := \log \text{vol}$, proving that $\mathcal{O}^+ := \lim_{t \rightarrow \beta} \mathcal{O}(t)$ exists and is regular. We may then apply the following, classical lemma to $N := M^{\text{reg}}$.

Lemma 3. *Let H be a smooth vector field on a manifold N .*

Let $p(t) : t \in (\alpha, \beta) \rightarrow N$ be a maximal integral curve and assume that there exists $q \in N$ such that $p(t) \rightarrow q$ as $t \rightarrow \beta$.

Then $\beta = \infty$ and $H(q) = 0$.

Notice also that $H = -\frac{1}{2}\nabla \log(\text{vol}^2) = -\frac{1}{2}\frac{\nabla \text{vol}^2}{\text{vol}^2}$. In the analytic context, this shows that if $\mathcal{O}_t \rightarrow \mathcal{O}$ and \mathcal{O} is singular, then $\|H_{\mathcal{O}_t}\| \rightarrow \infty$ (in the smooth category, following the idea of example 3, one could build examples for which such a limit does not exist). The following examples show that \mathcal{O} may be minimal or not.

Example 4. Let S^1 act on S^2 as in example 1 of section 2. Any \mathcal{O} that is not the equator or a pole flows, under MCF, to the closest pole, which is a singular, minimal, orbit. This happens in finite time.

Example 5. The above action of S^1 on S^2 induces an action of S^1 on $S^2 \times S^2$. Let $\mathcal{O} \simeq S^1 \times S^1$ be the product of a ‘‘small’’ orbit in S^2 (i.e., near a pole) and a ‘‘large’’ orbit in S^2 . The flow \mathcal{O}_t becomes singular as soon as the smaller orbit collapses onto the pole, but this limiting curve, $p \times S^1$, is not minimal: its flow exists until the second curve collapses.

Summarizing, we have proved the following result.

Theorem 2. *Let (M, g) be a compact, Riemannian manifold and let G be a compact group acting by isometries on M . Let \mathcal{O} be a principal orbit. Then:*

- (1) *MCF preserves orbits and types.*

Thus there exists a unique, maximal, curve of principal orbits \mathcal{O}_t , $t \in (\alpha, \beta)$, solution of MCF with $\mathcal{O}_0 = \mathcal{O}$.

- (2) *Assume (M, g) is analytic and $\mathcal{O}^+ := \lim_{t \rightarrow \beta} \mathcal{O}_t$ exists. Then:*

- *\mathcal{O}^+ is a regular orbit $\Leftrightarrow \beta = \infty \Leftrightarrow \|H(t)\| \rightarrow 0$.*

In this case, \mathcal{O}^+ is minimal.

- \mathcal{O}^+ is a singular orbit $\Leftrightarrow \beta < \infty \Leftrightarrow \|H(t)\| \rightarrow \infty$.
In this case, \mathcal{O}^+ may be minimal or not.
- If $\text{vol}(\mathcal{O}_t) \geq c > 0$, then \mathcal{O}^+ always exists and is regular.
- (3) Assume (M, g) is analytic. Then $\mathcal{O}^- := \lim_{t \rightarrow \alpha} \mathcal{O}_t$ always exists. It is a minimal regular orbit, $\alpha = -\infty$ and $\|H(t)\| \rightarrow 0$. In particular, “backwards MCF” always leads to a minimal regular orbit.

To get an analogous statement for flows of exceptional or singular orbits, it is sufficient to apply the theorem to the (smooth, compact) manifold M' defined as the closure in M of the set of orbits of the type in question: these orbits will be the principle orbits of the induced G -action on M' .

Remark. Using “equivariant Morse theory” applied to the volume function, it would be interesting to study the topology of Riemannian G -manifolds in terms of its minimal orbits. In theory, Theorem 2 would be useful in this.

4. LAGRANGIAN ORBITS AND MOMENT MAPS

We now want to focus on Lagrangian orbits generated by isometry groups of compact Kaehler manifolds. We start by recalling a few well-known facts concerning transformation groups of Riemannian and Kaehler manifolds. We refer to [K1] for proofs and further details.

Definition 1. Let (M, g) be a Riemannian manifold. A vector field X on M is an “infinitesimal isometry” if $\mathcal{L}_X g \equiv 0$; equivalently, if the local flow generated by X is a curve of isometries.

$\mathfrak{i}(M)$ will denote the space of all infinitesimal isometries. When (M, g) is complete, $\mathfrak{i}(M)$ is the Lie algebra of $\text{Isom}_g(M)$.

Definition 2. Let (M, J) be a complex manifold. A (real) vector field X on M is an “infinitesimal automorphism” if $\mathcal{L}_X J \equiv 0$; equivalently, if the local flow generated by X is a curve of automorphisms of (M, J) , or if $X - iJX$ is a holomorphic section of $T^{1,0}M$.

$\mathfrak{h}(M)$ will denote the set of infinitesimal automorphisms. It is closed wrt J and, when (M, J) is compact, it is the complex Lie algebra of the group $\text{Aut}_J(M)$ of automorphisms of (M, J) .

Theorem 3. Let (M, J, g, ω) be a compact Kaehler manifold. Then any infinitesimal isometry is an infinitesimal automorphism; so $\text{Isom}_g(M)^0 \leq \text{Aut}_J(M)^0$.

This, in turn, implies that $\text{Isom}_g(M)^0 \leq \text{Aut}_\omega(M)^0$.

The following proposition, although very simple, is the key to understanding Lagrangian orbits.

Proposition 3. Let (M^{2n}, J, g) be a compact Kaehler manifold and let $G \leq \text{Isom}_g(M)$ act on M with principal type (P) .

Assume there exists a regular Lagrangian G -orbit.

Then P is finite; so $\dim G = n$ and $\mathfrak{g}_p = \text{Lie}(G_p) = \{0\}$, $\forall p \in M^{\text{reg}}$.

Proof. Assume \mathcal{O} is a Lagrangian orbit. Then J gives an isomorphism $T\mathcal{O}^\perp \simeq T\mathcal{O} \simeq \mathfrak{g}/\mathfrak{p}$ that is equivariant wrt the natural P -action. Notice that, for each $p \in P$, this action coincides with the differential of the map

$$p : G/P \longrightarrow G/P, \quad p[g] := [pg] = [pgp^{-1}].$$

In other words, the action of p on $\mathfrak{g}/\mathfrak{p}$ is the differential of the adjoint action of p on G/P . Taken all together, these maps form a group homomorphism $P \longrightarrow GL(\mathfrak{g}/\mathfrak{p})$; the corresponding Lie algebra homomorphism is the map

$$\mathfrak{p} \longrightarrow gl(\mathfrak{g}/\mathfrak{p}), \quad X \mapsto [X, \cdot].$$

If \mathcal{O} is principal, the P -action on $T\mathcal{O}^\perp$ is trivial. Thus P acts trivially on $\mathfrak{g}/\mathfrak{p}$ (i.e., the action of each $p \in P$ on $\mathfrak{g}/\mathfrak{p}$ is the identity). So the map $\mathfrak{p} \longrightarrow gl(\mathfrak{g}/\mathfrak{p})$ is trivial (i.e., the action of each $X \in \mathfrak{p}$ is the zero map), i.e., \mathfrak{p} is an ideal of \mathfrak{g} . This implies that P^0 is normal in G and Corollary 2 of section 2 proves that P is finite.

Now assume \mathcal{O} is an exceptional Lagrangian orbit of type (K) . Locally, $M = G \times_K V$ and K acts as a finite group on $V = \mathfrak{g}/\mathfrak{k}$; so a neighborhood of $1 \in K$ acts trivially on $\mathfrak{g}/\mathfrak{k}$. This shows that K^0 acts trivially on $\mathfrak{g}/\mathfrak{k}$.

As above, K^0 is normal in G . Since $K^0 = P^0$, P^0 is also normal and we may conclude as above. \square

It is now convenient to introduce the concept of Hamiltonian group actions. Again, we refer to [A] for further details.

Let (M, ω) be a symplectic manifold. Recall that a vector field X on M is “Hamiltonian” if $\omega(X, \cdot)$ is an exact 1-form on M ; i.e., $\omega(X, \cdot) = df$, for some $f \in C^\infty(M)$. We say that f is a “Hamiltonian function” for X .

Definition 3. The action of G on M is “Hamiltonian” if the following conditions are satisfied:

- (1) There exists $\mu : M \longrightarrow \mathfrak{g}^*$ such that $\langle d\mu[p](\cdot), X \rangle = \omega[p](\tilde{X}, \cdot)$, where $\langle \cdot, \cdot \rangle$ denotes the natural pairing $\mathfrak{g}^* \times \mathfrak{g} \longrightarrow \mathbb{R}$. Equivalently, $\forall X \in \mathfrak{g}$, \tilde{X} is Hamiltonian (with Hamiltonian function $\mu_X := p \mapsto \langle \mu(p), X \rangle$).
- (2) μ is G -equivariant wrt the G -action on M and the co-adjoint G -action on \mathfrak{g}^* . Equivalently, $\mu_{[X, Y]}(p) = \omega[p](\tilde{X}, \tilde{Y})$.

We say that μ is a “moment map” for the action.

Remarks:

- (1) Assume (M, J, g, ω) is a compact Kaehler manifold and that, for some $G \leq Isom_g(M)^0$, condition (1) above is satisfied. Then, $\forall X \in \mathfrak{g}$,

$$d\mu_X = \omega(\tilde{X}, \cdot) = g(J\tilde{X}, \cdot).$$

This shows that $\nabla\mu_X = J\tilde{X}$, and so $\nabla\mu_X$ is an infinitesimal automorphism of (M, J) .

- (2) By definition, the differential $d\mu[p] : T_p M \longrightarrow \mathfrak{g}^*$ is the dual of the map

$$d\mu[p]^* : \mathfrak{g} \longrightarrow (T_p M)^*, \quad X \longrightarrow d\mu_X[p].$$

Thus $\text{Im } d\mu[p] = (\text{Ker } d\mu[p]^*)^\# = (\mathfrak{g}_p)^\#$, where $\mathfrak{g}_p = \text{Lie}(G_p)$.

In particular, $d\mu[p]$ is surjective iff G_p is discrete.

Definition 4. $\Sigma \subseteq (M, \omega)$ is “isotropic” if $\omega|_\Sigma \equiv 0$; if $\dim \Sigma = n$ and $\dim M = 2n$, then isotropic submanifolds are called “Lagrangian”.

We are mainly interested in moment maps for the following reason.

Lemma 4. Assume the action of G on (M, ω) is Hamiltonian, with moment map μ . Let $p \in M$. Then the following conditions are equivalent:

- (1) μ is constant on the orbit $\mathcal{O} = G \cdot p$.

- (2) \mathcal{O} is isotropic.
- (3) $\mu(p) \in [\mathfrak{g}, \mathfrak{g}]^\#$.

We now have all the elements necessary to prove the following:

Corollary 5. *Let (M, J, g) be a compact Kaehler manifold. Assume that $G \leq \text{Isom}_g(M)$ acts in a Hamiltonian fashion. Then the set*

$$\mathcal{L}(M; G) := \{p \in M^{\text{reg}} : G \cdot p \text{ is a Lagrangian orbit}\}$$

either is empty or is a smooth submanifold of M^{reg} , of dimension $2n - \dim[\mathfrak{g}, \mathfrak{g}]$.

Proof. If there exists a regular Lagrangian orbit, then, by Proposition 3, P is finite and $\dim G = n$. Thus every regular isotropic orbit has dimension n and is Lagrangian. Lemma 4 now shows that, if we let μ_{reg} denote the restriction of μ to M^{reg} , then $\mathcal{L}(M; G) = \mu_{\text{reg}}^{-1}([\mathfrak{g}, \mathfrak{g}]^\#)$. Since $\mathfrak{g}_p = 0$, μ_{reg} is a submersion; so $\mathcal{L}(M, G)$ is smooth, of dimension $n + \dim [\mathfrak{g}, \mathfrak{g}]^\#$. \square

Example 6. Assume $G \leq \text{Isom}_g(M)$ is semisimple. Then the G -action on M is Hamiltonian (cf. [A]) and Lagrangian orbits are isolated.

Example 7. Assume that a torus $T^n \leq \text{Isom}_g(M)$ acts effectively on M and that $H^1(M; \mathbb{R}) = 0$. Then the action is Hamiltonian (cf. [A]), $P = 1$, $[\mathfrak{g}, \mathfrak{g}] = 0$ and every regular orbit is Lagrangian. In other words, $\mathcal{L}(M, G) = M^{\text{reg}}$.

In particular, there exists a minimal, Lagrangian orbit (cf. also [G]).

An example of this is provided by the standard T^n -action on \mathbb{P}^n .

5. MCF OF LAGRANGIAN ORBITS IN KE MANIFOLDS

In this section, we will assume that (M, J, g, ω) is a KE manifold.

Since we are interested in group actions, we must recall (cf. [K1]) some basic facts concerning their transformation groups.

Theorem 4. *Let M be a compact KE manifold such that $\text{Ric} = c \cdot g$, $c > 0$.*

For any (real) vector field X , let $Z := X - iJX$ and $\zeta := g(Z, \cdot)$. Then

- (1) $i(M)$ is totally real in $\mathfrak{h}(M)$: i.e., if $X \in i(M)$, then $JX \notin i(M)$.
- (2) $X \in \mathfrak{h}(M) \Leftrightarrow \zeta = \bar{\partial}f : f \in C^\infty(M; \mathbb{C}), \Delta f = 2cf$.

In particular, $\int_M f = 0$; so such an f is unique.

- (3) $X \in i(M) \Leftrightarrow \text{Re}(f) = 0$.

If we set $E_{2c} := \{f \in C^\infty(M; \mathbb{R}) : \Delta f = 2cf\}$, there is an isomorphism:

$$E_{2c} \simeq i(M), \quad f \mapsto i\bar{\partial}f = \zeta.$$

- (4) $\mathfrak{h}(M) = i(M) \oplus Ji(M)$.

It is possible (cf. [K2]) to prove that positive compact KE manifolds are simply connected. This implies that every fundamental vector field induced by $G \leq \text{Isom}_g(M)$ is Hamiltonian. From our point of view, however, much more is true:

Proposition 4 (cf. [F]). *Let M be a compact positive KE manifold and $G \leq \text{Isom}_g(M)$. Then the action of G on M is Hamiltonian.*

Recall the correspondence and the notation from Theorem 4 above:

$$X \in i(M) \leftrightarrow f : f \in C^\infty(M; \mathbb{R}), \quad \Delta f = 2cf, \quad \zeta = i\bar{\partial}f.$$

Then $\mu_X := -\frac{1}{2}f$ defines a moment map for M, G .

Moment maps are usually not unique: if μ is a moment map, $\mu + c$ also is, for any $c \in [\mathfrak{g}, \mathfrak{g}]^\# \leq \mathfrak{g}^*$. The proposition above suggests the following.

Definition 5. The moment map defined in Proposition 4 above will be called the “canonical moment map” of the G -action.

Recall, however, that moment maps are uniquely defined on $[\mathfrak{g}, \mathfrak{g}]$ because $\mu_{[X,Y]} = \omega(\tilde{X}, \tilde{Y})$. Recall also Lemma 4. Proposition 4 thus leads to the following result:

Corollary 6. Let M be a compact KE manifold such that $\text{Ric} = c \cdot g, c > 0$.

- (1) $\forall X, Y \in \mathfrak{i}(M), \omega(\tilde{X}, \tilde{Y}) \in E_{2c}$.
- (2) $\forall f \in E_{2c}, f$ restricted to $\mathcal{L}(M; G)$ is G -invariant.

Putting everything together and using the fact that KE metrics are analytic, we can now prove the following result:

Theorem 5. Let M be a compact positive KE manifold and let $G \leq \text{Isom}_g(M)$.

Assume $\mathcal{L}(M; G)$ is not empty. Then H is tangent to $\mathcal{L}(M; G)$. Thus MCF preserves the Lagrangian condition and may be studied as in Theorem 2.

Furthermore, there exists a minimal Lagrangian orbit.

Proof. Recall, for any Lagrangian submanifold Σ immersed in Kaehler M , the isomorphism

$$(T\Sigma)^\perp \simeq \Lambda^1(\Sigma), \quad V \simeq \nu := \omega(V, \cdot)|_\Sigma.$$

It is well known (cf. [TY]) that if $\sigma_H \in \Lambda^1(\Sigma)$ denotes the 1-form corresponding to the mean curvature vector field H under this isomorphism, then $d\sigma_H = \rho|_\Sigma$, where $\rho(X, Y) := \text{Ric}(JX, Y)$ is the “Ricci 2-form” of M .

When M is KE, $\rho = c \cdot \omega$; so this shows that σ_H is closed.

Now let $p \in \mathcal{L}(M; G)$. Then $T_p\mathcal{L} = \{X \in T_pM : d\mu[p](X) \in [\mathfrak{g}, \mathfrak{g}]^\#\}$. So we need to prove that $d\mu[p](H) \in [\mathfrak{g}, \mathfrak{g}]^\#$.

Since H is G -invariant, σ_H also is; i.e., $\sigma_H \in \mathfrak{g}^*$. Notice that $d\mu[p](H) = \omega[p](\cdot, H) = -\sigma_H[p]$.

Recall that, for any 1-form $\alpha \in \Lambda^1(\Sigma)$,

$$d\alpha(X, Y) = X\alpha(Y) - Y\alpha(X) - \alpha[X, Y].$$

Thus $0 = d\sigma_H(X, Y) = -\sigma_H[X, Y], \forall X, Y \in \mathfrak{g}$, as desired.

The first claim is now obvious. The properties of vol show that there is a Lagrangian orbit \mathcal{O} of maximum volume (which is minimal in $\mathcal{L}(M; G)$). Let \mathcal{O}_t be a curve in $\mathcal{L}(M; G)$ such that $\mathcal{O}_0 = \mathcal{O}$ and $\frac{d}{dt}\mathcal{O}_t = H$. Then $0 = \frac{d}{dt}\text{vol}(\mathcal{O}_t)|_{t=0} = -\int_M (H\mathcal{O}, H\mathcal{O});$ so $H\mathcal{O} \equiv 0$. \square

Remark. When M is compact Kaehler Ricci-flat, one can show that $\text{Isom}_g(M)^0$ is a torus. So example 7 shows that the analogous statement is trivially true.

When M is compact negative KE, $\text{Isom}_g(M)^0 = \{Id\}$. So these manifolds are not interesting from our point of view. Cf. [K1] for details.

Our final goal is to explore the relationship between MCF and the canonical moment map.

Proposition 5. Let M^{2n} be a compact KE manifold such that $\text{Ric} = c \cdot g, c > 0$. Given $G \leq \text{Isom}_g(M)$, assume that $\mathcal{L}(M; G)$ is not empty. Let $\mu : M \rightarrow \mathfrak{g}^*$ denote the canonical moment map. Then, on $\mathcal{L}(M; G)$,

- (1) $\forall X \in \mathfrak{g}, \quad H\mathcal{O} \cdot \nabla\mu_X = c\mu_X$.

- (2) $\forall p \in \mathcal{L}(M; G)$, the natural (G -invariant) metric on $G \cdot p \subseteq M$ defines metrics on \mathfrak{g} and \mathfrak{g}^* . For the induced norm (which depends on p),

$$d\|\mu\|^2[p](H) = 2c\|\mu(p)\|^2.$$

Proof. Let $\mathcal{O} \simeq G \cdot p/G_p$ denote any regular Lagrangian orbit. Let $e_1, \dots, e_n \in \mathfrak{g} \simeq T_p\mathcal{O}$ be an orthonormal basis wrt the induced metric. To simplify the notation, we will denote the corresponding fundamental vector fields also by e_i . Then

$$(H_{\mathcal{O}}, \nabla\mu_X) = (\nabla_{e_j}^\perp e_j, \nabla\mu_X) = (\nabla_{e_j} e_j, \nabla\mu_X) = -(e_j, \nabla_{e_j} \nabla\mu_X).$$

In section 4, we saw that $\nabla\mu_X$ is an infinitesimal automorphism of M . Thus

$$\nabla_{Je_j} \nabla\mu_X = \nabla_{\nabla\mu_X} Je_j + [Je_j, \nabla\mu_X] = J(\nabla_{\nabla\mu_X} e_j + [e_j, \nabla\mu_X]) = J(\nabla_{e_j} \nabla\mu_X).$$

The definition of the canonical moment map now shows that

$$\begin{aligned} 2(H_{\mathcal{O}}, \nabla\mu_X) &= -(e_j, \nabla_{e_j} \nabla\mu_X) - (Je_j, J(\nabla_{e_j} \nabla\mu_X)) \\ &= -(e_j, \nabla_{e_j} \nabla\mu_X) - (Je_j, \nabla_{Je_j} \nabla\mu_X) \\ &= -\operatorname{div}_M(\nabla\mu_X) = \Delta_M \mu_X = 2c\mu_X. \end{aligned}$$

This proves (1). Applying (1) to $X = e_i$, multiplying by $2\mu_{e_i}$ and summing wrt i shows that

$$H_{\mathcal{O}} \cdot \nabla\|\mu\|^2 = 2c\|\mu\|^2,$$

which is (2). □

We can now prove

Theorem 6. *Let M^{2n} be a compact KE manifold such that $\operatorname{Ric} = c \cdot g, c > 0$. For $G \leq \operatorname{Isom}_g(M)$, let μ denote the canonical moment map and let $E_{2c}(G) := \{f \in E_{2c} : f = \mu_X, \text{ for some } X \in \mathfrak{g}\}$.*

Assume that regular orbits have dimension n . Then a Lagrangian orbit \mathcal{O} is minimal iff $\mu(\mathcal{O}) = 0$. In particular, minimal Lagrangian orbits are isolated. Furthermore, the following are equivalent:

- (1) *There exists a Lagrangian orbit.*
- (2) *There exists a minimal Lagrangian orbit.*
- (3) $0 \in \mu(M)$.
- (4) *The set $\{p \in M : f(p) = 0, \forall f \in E_{2c}(G)\}$ is not empty.*

Proof. By hypothesis, an orbit is regular iff it is n -dimensional. In particular, every Lagrangian orbit \mathcal{O} is regular. We may thus restrict our attention to M^{reg} .

If \mathcal{O} is minimal, Proposition 5 shows that $\mu(\mathcal{O}) = 0$. Vice versa, assume that $\mu(\mathcal{O}) = 0$. Let \mathcal{O}_t be obtained by MCF applied to \mathcal{O} . Then Proposition 5 shows that $f(t) := \|\mu\|^2(\mathcal{O}_t)$ satisfies

$$\frac{d}{dt} f(t) = 2cf, \quad f(0) = 0.$$

This implies that $f(t) \equiv 0$; so $\mathcal{O}(t) \subseteq \mu^{-1}(0)$. However, μ is a submersion; so $\mu^{-1}(0)$ is smooth of dimension n and, since P is finite, the elements of $\mu^{-1}(0)/G$ are isolated. Thus $\mathcal{O}(t) \equiv \mathcal{O}$, i.e., \mathcal{O} is minimal.

Together with Theorem 5, this shows that (1), (2) and (3) are equivalent. The equivalence of (3) and (4) comes directly from the definition of μ . □

Remark. In the toric case, one can show that $\mu^{-1}(0)$ is connected. So Theorem 6 implies that the minimal Lagrangian orbit is unique. This result was obtained also in [G], by lifting the T^n -action from M to its canonical bundle K_M and studying the induced geometry.

ACKNOWLEDGEMENTS

I wish to thank T. Ilmanen for a useful conversation and P. de Bartolomeis and G. Tian for their long-term support, suggestions and interest. I also gratefully acknowledge the generous support of the University of Pisa and of GNSAGA, and the hospitality of MIT.

REFERENCES

- [A] Audin, M., The topology of torus actions on symplectic manifolds, Progress in Math., vol. 93, Birkhäuser-Verlag, Basel, 1991 MR **92m**:57046
- [F] Futaki, A., The Ricci curvature of symplectic quotients of Fano manifolds, Tohoku Math. J., 39 (1987), 329-339 MR **88m**:53124
- [G] Goldstein, E., Minimal Lagrangian tori in Kaehler Einstein manifolds, math.DG/0007135 (preprint)
- [H] Hsiang, W., On the compact homogeneous minimal submanifolds, Proc. Nat. Acad. Sci., 56 (1966), pp. 5-6 MR **34**:5037
- [HL] Hsiang, W. and Lawson, H. B., Jr., Minimal submanifolds of low cohomogeneity, J. Differential Geometry, 5 (1971), pp. 1-38 MR **45**:7645
- [K1] Kobayashi, S., Transformation groups in differential geometry, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 70, Springer-Verlag, 1972 MR **50**:8360
- [K2] Kobayashi, S., On compact Kaehler manifolds with positive definite Ricci tensor, Ann. of Math. (2), 74 (1961), pp. 570-574 MR **24**:A2922
- [TY] Thomas, R. and Yau, S.-T., Special Lagrangians, stable bundles and mean curvature flow, math.DG/0104197 (preprint)

IMPERIAL COLLEGE, LONDON, UK

UNIVERSITY OF PISA, PISA, ITALY

E-mail address: `pacini@paley.dm.unipi.it`

Current address: Department of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332

E-mail address: `pacini@math.gatech.edu`

SINGULAR INTEGRALS ON SYMMETRIC SPACES, II

ALEXANDRU D. IONESCU

ABSTRACT. We extend some of our earlier results on boundedness of singular integrals on symmetric spaces of real rank one to arbitrary noncompact symmetric spaces. Our main theorem is a transference principle for operators defined by \mathbb{K} -bi-invariant kernels with certain large scale cancellation properties. As an application we prove L^p boundedness of operators defined by Fourier multipliers that satisfy singular differential inequalities of the Hörmander–Michlin type.

1. INTRODUCTION

Let \mathbb{G} be a noncompact connected semisimple Lie group with finite center, \mathbb{K} a maximal compact subgroup and $\mathbb{X} = \mathbb{G}/\mathbb{K}$ an associated symmetric space. In this paper we study L^p boundedness properties of a class of operators on the symmetric space \mathbb{X} which are the analogue of the singular integral operators on Euclidean spaces. By analogy with the Euclidean case these operators can be defined by Fourier multipliers or by convolution with Calderon–Zygmund kernels.

A necessary condition for L^p boundedness of an operator T_m defined by a multiplier m is that the multiplier extends to a bounded W -invariant holomorphic function in the interior of the tube $\mathcal{T}_p = \mathfrak{a}^* + i\text{co}(W \cdot \rho_p)$, where $\rho_p = |2/p - 1|\rho$. This was observed by Clerc and Stein [5], who also proved a sufficient condition for L^p boundedness when the group \mathbb{G} is complex. By analogy with the Euclidean case, a natural theorem is the following: assume that $p \in (1, 2) \cup (2, \infty)$ and the multiplier m extends to a holomorphic function in the interior of the tube \mathcal{T}_p . Assume in addition that m satisfies differential inequalities of the form

$$(1.1) \quad \left| \partial_\lambda^j m(\lambda) \right| \leq A_j (1 + |\lambda|)^{-j}$$

for all $\lambda \in \mathcal{T}_p$ and any $j \geq 0$, where $\partial_\lambda^j m(\lambda)$ denotes any partial derivative of m of order j . Then the operator T_m extends to a bounded operator on $L^p(\mathbb{X})$. Statements of this type in various settings can be found in [5], [18], [4], [22], [1], [8] and [9].

In this paper we deal with operators defined by multipliers that are allowed to have singularities on the boundary of the tube \mathcal{T}_p . These multipliers satisfy differential inequalities of the form

$$\left| \partial_\lambda^j m(\lambda) \right| \leq A_j d(\lambda)^{-j}$$

Received by the editors September 12, 2001.

2000 *Mathematics Subject Classification.* Primary 22E46, 43A85.

The author was supported in part by the National Science Foundation under NSF Grant No. 0100021.

where the function $d(\lambda)$ can vanish on the boundary of the tube \mathcal{T}_p (as in (4.1)). These differential inequalities are the analog on symmetric spaces of the classical Hörmander–Michlin differential inequalities. Natural examples are provided by imaginary powers of a suitably modified Laplacian, as in [8]. The associated multipliers have singularities at the points $\{w \cdot (i\rho_p) : w \in W\}$.

The main difficulty in proving L^p boundedness of operators defined by multipliers with singularities on the boundary is that the classical Herz majorizing principle [13] fails to apply. As in the case of Euclidean spaces, it is critical to be able to use the large scale cancellation properties of the kernel of such an operator. Our main theorem in this direction is Theorem 3.1—a transference principle that allows one to deal with convolution operators defined by kernels of the form

$$(1.2) \quad K(\exp H) = e^{-2\rho(H)/p} \phi(H),$$

where ϕ is a Calderon–Zygmund kernel on \mathfrak{a} . Here $1 < p < 2$ and the factor $e^{-2\rho(H)/p}$ is the critical exponential decay factor for L^p boundedness. The proof of this theorem is similar to the proof of the corresponding theorem on symmetric spaces of real rank one ([14, Theorem 1]).

Our second main task is to show that this transference principle, together with the Herz majorizing principle and local analysis, suffices to prove L^p boundedness of operators defined by multipliers satisfying (4.1). This is significantly harder in the case of general symmetric spaces than in the case of real rank one symmetric spaces. One of the difficulties arises from the fact that the Harish-Chandra expansion of the elementary spherical functions does not converge uniformly in the region close to the walls of the positive Weyl chamber. Thus one cannot get good pointwise estimates on the kernels of the operators in this region. This difficulty was overcome by Anker [1] and we follow his approach. The main observation we need to make is that the volume of this region is smaller than the volume of the full Weyl chamber (in the sense of (4.6)). We will then apply our transference theorem to the part of the kernel supported away from the walls of the positive Weyl chamber. The Harish-Chandra expansion converges uniformly in this region and the kernel corresponding to the main term in this expansion is of the form (1.2). Also, one can use the Herz majorizing principle to deal with the error terms in the Harish-Chandra expansion in this region.

2. PRELIMINARIES

Most of our notation related to semisimple Lie groups and symmetric spaces is standard and can be found, for example, in [12]. Let G be a noncompact connected semisimple Lie group with finite center, \mathfrak{g} the Lie algebra of G , θ a Cartan involution of \mathfrak{g} and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ the associated Cartan decomposition. Let $K = \exp \mathfrak{k}$ be a maximal compact subgroup of G and let $X = G/K$ be an associated symmetric space. Let \mathfrak{a} be a maximal abelian subspace of \mathfrak{p} , M the centralizer of $\exp \mathfrak{a}$ in K , Σ the restricted root system of the pair $(\mathfrak{g}, \mathfrak{a})$ and W the associated Weyl group. Let $\mathfrak{a}^+ \subset \mathfrak{a}$ be a positive Weyl chamber and Σ^+ the corresponding set of positive roots. For any root $\alpha \in \Sigma$ let \mathfrak{g}_α be the root space associated to α and m_α its dimension. Let $\mathfrak{n} = \sum_{\alpha \in \Sigma^+} \mathfrak{g}_\alpha$, $\bar{\mathfrak{n}} = \theta(\mathfrak{n})$, $N = \exp \mathfrak{n}$ and $\bar{N} = \exp \bar{\mathfrak{n}}$.

The group G has an Iwasawa decomposition $G = K(\exp \mathfrak{a})N$ and a Cartan decomposition $G = K(\exp \overline{\mathfrak{a}^+})K$. For each $g \in G$ denote by $H(g) \in \mathfrak{a}$ and $g^+ \in \overline{\mathfrak{a}^+}$ the middle components of g in these decompositions. We will also use the Iwasawa

decomposition $\mathbb{G} = \overline{\mathbb{N}}(\exp \mathfrak{a})\mathbb{K}$. We normalize the Haar measure dk on \mathbb{K} with total mass 1 and the Haar measure $d\overline{n}$ on $\overline{\mathbb{N}}$ such that

$$\int_{\overline{\mathbb{N}}} e^{-2\rho(H(\overline{n}))} d\overline{n} = 1,$$

where

$$\rho(H) = \frac{1}{2} \sum_{\alpha \in \Sigma^+} m_\alpha \alpha(H).$$

The Haar measure dg on \mathbb{G} can be normalized so that

$$\int_{\mathbb{G}} F(g) dg = \int_{\overline{\mathbb{N}}} \int_{\mathfrak{a}} \int_{\mathbb{K}} F(\overline{n}(\exp H)k) e^{2\rho(H)} dk dH d\overline{n}$$

and

$$\int_{\mathbb{G}} F(g) dg = \int_{\mathbb{K}} \int_{\mathbb{K}} \int_{\mathfrak{a}_+} F(k_1(\exp H)k_2) \delta(H) dH dk_1 dk_2$$

for any continuous compactly supported function $F : \mathbb{G} \rightarrow \mathbb{C}$. One has by definition

$$(2.1) \quad \delta(H) = C \prod_{\alpha \in \Sigma^+} (\sinh \alpha(H))^{m_\alpha}.$$

The Killing form of \mathfrak{g} induces a scalar product $\langle \cdot, \cdot \rangle$ on \mathfrak{a} . Denote by \mathfrak{a}^* and $\mathfrak{a}_{\mathbb{C}}^*$ the real and the complex dual of \mathfrak{a} . For any $\lambda \in \mathfrak{a}^*$ let H_λ be the unique element of \mathfrak{a} with the property that $\langle H_\lambda, H \rangle = \lambda(H)$ for any $H \in \mathfrak{a}$. We transfer the scalar product $\langle \cdot, \cdot \rangle$ to \mathfrak{a}^* and extend it to a \mathbb{C} -bilinear form on $\mathfrak{a}_{\mathbb{C}}^*$. The Fourier transform of a smooth compactly supported function $f : \mathbb{X} \rightarrow \mathbb{C}$ is by definition the function $\tilde{f} : \mathfrak{a}_{\mathbb{C}}^* \times \mathbb{K}/\mathbb{M} \rightarrow \mathbb{C}$ given by

$$(2.2) \quad \tilde{f}(\lambda, b) = \int_{\mathbb{X}} f(z) e^{(-i\lambda + \rho)A(z, b)} dz$$

where $A(z, b)$ is an \mathfrak{a} -valued analogue of the usual scalar product on Euclidean spaces (see [12, Chapter III]). For any $g \in \mathbb{G}$ and $k \in \mathbb{K}$ one has by definition $A(g\mathbb{K}, k\mathbb{M}) = -H(g^{-1}k)$. If f is \mathbb{K} -invariant (i.e., $f(k \cdot z) = f(z)$ for any $k \in \mathbb{K}$ and $z \in \mathbb{X}$), then \tilde{f} does not depend on b . The formula (2.2) becomes

$$(2.3) \quad \tilde{f}(\lambda, b) = \int_{\mathbb{G}} f(g\mathbb{K}) \varphi_{-\lambda}(g) dg$$

where

$$(2.4) \quad \varphi_\lambda(g) = \int_{\mathbb{K}/\mathbb{M}} e^{(i\lambda + \rho)A(g\mathbb{K}, b)} db.$$

A central result in the theory of convolution operators on semisimple Lie groups is the Kunze–Stein phenomenon, which states that if $p \in [1, 2)$, $f \in L^2(\mathbb{G})$ and $K \in L^p(\mathbb{G})$, then

$$(2.5) \quad \|f * K\|_{L^2(\mathbb{G})} \leq C_p \|f\|_{L^2(\mathbb{G})} \|K\|_{L^p(\mathbb{G})}.$$

This inequality was proved by Kunze and Stein [15] in the case when the group \mathbb{G} is $\mathrm{SL}(2, \mathbb{R})$ (and, later on, for other particular groups) and by Cowling [7] in the general case. In [13] Herz noticed that the inequality (2.5) can be sharpened (and its proof greatly simplified) if the kernel K is \mathbb{K} -bi-invariant, i.e., $K(k_1 g k_2) = K(g)$ for any $k_1, k_2 \in \mathbb{K}$ and $g \in \mathbb{G}$. Let $\|f * K\|_{L^p(\mathbb{G})}$ denote the norm of the convolution operator defined by the kernel K and for any $p \in [1, \infty]$ let $\rho_p = |2/p - 1|\rho$. One has the following criterion due to Herz [13]:

Proposition 2.1. *If $1 \leq p \leq \infty$ and K is a \mathbb{K} -bi-invariant kernel on \mathbb{G} , then*

(2.6)
$$||| * K |||_{L^p(\mathbb{G})} \leq C \int_{\mathbb{G}} |K(g)| \varphi_{-i\rho_p}(g) dg.$$

It is easy to see that if $1 \leq p < 2$ and $H \in \overline{\mathfrak{a}^+}$ (the closure of the cone \mathfrak{a}^+), then $\varphi_{-i\rho_p}(\exp H) \leq C_p e^{-2\rho(H)/p'}$, where $p' = p/(p - 1)$ is the conjugate exponent of p . Thus (2.6) becomes

(2.7)
$$||| * K |||_{L^p(\mathbb{G})} \leq C_p \int_{\overline{\mathfrak{a}^+}} |K(\exp H)| \delta(H) e^{-2\rho(H)/p'} dH.$$

3. A TRANSFERENCE THEOREM

Let $\phi : \mathfrak{a} \rightarrow \mathbb{C}$ be a function on \mathfrak{a} supported in $\overline{\mathfrak{a}^+}$. Let $\ell = \dim \mathfrak{a}$ denote the rank of the group \mathbb{G} . Assume that the function ϕ satisfies the following basic assumption: there exist two constants A and $c_0 > 0$ such that

(3.1)
$$|\phi(H_1) - \phi(H_2)| \leq A(1 + |H_1|)^{-(\ell + c_0)}$$

for any $H_1, H_2 \in \overline{\mathfrak{a}^+}$ with the property that $|H_1 - H_2| \leq 1$.

Let p be a fixed exponent in the interval $(1, 2)$ and let $K_{p,\phi} : \mathbb{G} \rightarrow \mathbb{C}$ be the \mathbb{K} -bi-invariant kernel on \mathbb{G} given by

(3.2)
$$K_{p,\phi}(k_1(\exp H)k_2) = e^{-2\rho(H)/p} \phi(H)$$

for any $H \in \overline{\mathfrak{a}^+}$ and $k_1, k_2 \in \mathbb{K}$. As in the previous section, let

$$||| * K_{p,\phi} |||_{L^p(\mathbb{X})} = ||| * K_{p,\phi} |||_{L^p(\mathbb{G})}$$

denote the norm of the convolution operator with the kernel $K_{p,\phi}$ on $L^p(\mathbb{X})$ and $||| * \phi |||_{L^p(\mathfrak{a})}$ the norm of the convolution operator with the kernel ϕ on $L^p(\mathfrak{a})$. Our first main theorem is the following:

Theorem 3.1. *Assume that ϕ satisfies (3.1), $p \in (1, 2)$, and $K_{p,\phi}$ is defined as in (3.2). Then there exists a constant C_p depending only on p and the group \mathbb{G} such that*

$$||| * K_{p,\phi} |||_{L^p(\mathbb{X})} \leq C_p \left(||| * \phi |||_{L^p(\mathfrak{a})} + \tilde{A} \right).$$

The constant \tilde{A} depends only on ℓ and the constants A and c_0 in (3.1).

Remark 1. This theorem is sharper than the classical transference principle of Coifman and Weiss [6, Theorem 8.7] because the factor that makes the transition between the kernels ϕ and $K_{p,\phi}$ is $e^{2\rho(H)/p}$. This transition factor is $\delta(H)$ in [6, Theorem 8.7], which is proportional to $e^{2\rho(H)}$ if H is not close to the walls of the Weyl chamber \mathfrak{a}^+ . See also [8, Theorem 4.1] and [9] for other transference principles.

As an application, we have a new variant of Herz’s majorizing principle. Assume that the function ϕ is supported in the cone $\overline{\mathfrak{a}^+}$ and satisfies the differential inequalities

(3.3)
$$|\partial_H^j \phi(H)| \leq A'(1 + |H|)^{-\ell - j}$$

for any $H \in \mathfrak{a}^+$ and $j = 0$ or $j = 1$, where $\partial_H^j \phi(H)$ denotes any partial derivative of ϕ of order j . Assume also that the function ϕ satisfies the cancellation condition

(3.4)
$$\left| \int_{|H| \leq N} \phi(H) dH \right| \leq A'$$

for any $N > 0$. In this case the function ϕ is a kernel of the Calderon–Zygmund type on \mathfrak{a} . Indeed, the conditions (3.3) and (3.4) guarantee the fact that the kernel ϕ defines a bounded operator on $L^q(\mathfrak{a})$ for any $q \in (1, \infty)$ (see, for example, [20, Chapters I and VI]). One has the following consequence of Theorem 3.1:

Corollary 3.2. *If $p \in (1, 2)$ and ϕ satisfies (3.3) and (3.4), then*

$$||| * K_{p,\phi} |||_{L^p(\mathbb{G})} \leq C_p.$$

The constant C_p depends only on p and A' .

Notice that, in general, the function ϕ in Corollary 3.2 is not absolutely integrable at ∞ ; therefore, this corollary cannot be obtained as a consequence of the Kunze–Stein phenomenon or Herz majorizing principle. Easy examples, similar to the ones in [14], show that the large scale cancellation condition (3.4) is crucial. We also remark that both Theorem 3.1 and Corollary 3.2 are false if $p = 2$.

The rest of this section is devoted to proving Theorem 3.1. We will follow closely the line of the proof of Theorem 1 in [14]. Throughout the proof of Theorem 3.1 the letter C will denote universal constants depending only on the group \mathbb{G} , C_p will denote constants depending on p and the group \mathbb{G} , and \tilde{A} will denote constants depending only on ℓ and the values of c_0 and A in (3.1). Elements of \mathfrak{a} will be denoted by H , H_1 , and H_2 , and elements of $\bar{\mathbb{N}}$ will be denoted by \bar{n} , \bar{m} , and \bar{v} .

Recall that for any locally integrable \mathbb{K} -bi-invariant function K and any smooth compactly supported function $f : \mathbb{X} \rightarrow \mathbb{C}$ the convolution $f * K$ is defined by the formula

$$f * K(z) = \int_{\mathbb{G}} f(h \cdot \mathbf{0}) K(h^{-1} \cdot z) dh,$$

where $\mathbf{0} = \mathbb{G}/\mathbb{K}$ is the origin of the symmetric space \mathbb{X} . Notice that

$$||| * K_{p,\phi} |||_{L^p(\mathbb{X})} = \sup_{\|f\|_p = \|g\|_{p'} = 1} \left| \int_{\mathbb{G}} \int_{\mathbb{G}} f(h \cdot \mathbf{0}) K_{p,\phi}(h^{-1}h') g(h' \cdot \mathbf{0}) dh dh' \right|,$$

where the supremum is taken over smooth compactly supported functions $f, g : \mathbb{X} \rightarrow \mathbb{C}$, and $p' = p/(p-1)$ is the conjugate exponent of p .

As in the proof of Theorem 1 in [14], the main idea is to estimate integrals in Iwasawa coordinates. We identify the group \mathbb{G} with $\bar{\mathbb{N}} \times \mathfrak{a} \times \mathbb{K}$ using the Iwasawa decomposition $\mathbb{G} = \bar{\mathbb{N}}(\exp \mathfrak{a})\mathbb{K}$. Let $d\mu$ be the measure on $\bar{\mathbb{N}} \times \mathfrak{a}$ corresponding to this decomposition, i.e., $d\mu = e^{2\rho(H)} d\bar{n} dH$. Thus it suffices to prove that for any smooth compactly supported functions $f, g : \bar{\mathbb{N}} \times \mathfrak{a} \rightarrow \mathbb{C}$ one has

$$(3.5) \quad |I_{p,\phi}(f, g)| \leq C_p (||| * \phi |||_{L^p(\mathfrak{a})} + \tilde{A}) \|f\|_{L^p(\bar{\mathbb{N}} \times \mathfrak{a}, d\mu)} \|g\|_{L^{p'}(\bar{\mathbb{N}} \times \mathfrak{a}, d\mu)}$$

where

$$(3.6) \quad I_{p,\phi}(f, g) = \int_{\bar{\mathbb{N}}} \int_{\bar{\mathbb{N}}} \int_{\mathfrak{a}} \int_{\mathfrak{a}} f(\bar{m}, H_1) K_{p,\phi}(\delta_{-H_1}(\bar{m}^{-1}\bar{n})(\exp(H_2 - H_1))) \\ \times g(\bar{n}, H_2) e^{2\rho(H_1 + H_2)} dH_1 dH_2 d\bar{m} d\bar{n}.$$

By definition $\delta_H(\bar{v}) = (\exp H)\bar{v}(\exp -H)$ for any $H \in \mathfrak{a}$ and $\bar{v} \in \bar{\mathbb{N}}$. It is clear that δ_H is a dilation of the group $\bar{\mathbb{N}}$. In order to estimate $|I_{p,\phi}(f, g)|$ we need to understand the connection between the Iwasawa decomposition and the Cartan decomposition of the group \mathbb{G} . Let ${}^+\mathfrak{a}$ denote the interior of the cone $\{H \in \mathfrak{a} :$

$\langle H, H' \rangle > 0$ for any $H' \in \mathfrak{a}^+$. It is well known that $\mathfrak{a}^+ \subset {}^+\mathfrak{a}$ (see [10, Lemma 35]). For any $H \in \mathfrak{a}^+$ let

$$\omega(H) = \min_{\alpha \in \Sigma^+} \alpha(H).$$

Lemma 3.3. *If $\bar{v} \in \overline{\mathfrak{N}}$ and $H \in \mathfrak{a}^+$, then*

$$(3.7) \qquad [\bar{v}(\exp H)]^+ = H + H(\bar{v}) + H'(\bar{v}, H)$$

where $H'(\bar{v}, H) \in \overline{{}^+\mathfrak{a}}$ and there exists a constant C_0 such that

$$(3.8) \qquad e^{2\rho(H'(\bar{v}, H))} - 1 \leq C e^{-2\omega(H)} e^{C_0\rho(H(\bar{v}))}.$$

Proof of Lemma 3.3. This lemma is a quantitative version of Proposition 4.24 in [12, Chapter II]. Recall that g^+ and $H(g)$ denote the \mathfrak{a} -components of the element $g \in \mathbb{G}$ in the Cartan decomposition and the Iwasawa decomposition of the group \mathbb{G} . It is well known that $g^+ - H(g) \in \overline{{}^+\mathfrak{a}}$ for any $g \in \mathbb{G}$ ([10, Lemma 35]); therefore $H'(\bar{v}, H) \in \overline{{}^+\mathfrak{a}}$. The estimate (3.8) is a consequence of the proof of Proposition 4.24 in [12, Chapter II]. With the notation in this proposition, let π be an irreducible representation of \mathbb{G} on a finite-dimensional vector space V , Λ the highest weight of π , and ν the restriction of Λ to \mathfrak{a} . Let

$$V = V_\nu + \sum_{\mu < \nu} V_\mu = V_\nu + W$$

be the decomposition of V into weight subspaces. Fix v_1, \dots, v_r and w_1, \dots, w_p , the corresponding orthogonal basis consisting of restricted weight vectors. It follows immediately from one of the equations in Proposition 4.24 in [12, Chapter II] that

$$r(e^{2\nu(H'(\bar{v}, H))} - 1) \leq \sum_{j=1}^p \|\pi(\bar{v})w_j\|^2 e^{(2\mu_j - 2\nu)(H) - 2\nu(H(\bar{v}))}.$$

One clearly has

$$\|\pi(\bar{v})w_j\| \leq e^{\nu(\bar{v}^+)}.$$

Therefore, it suffices to prove that for any $\bar{v} \in \overline{\mathfrak{N}}$,

$$(3.9) \qquad \rho(\bar{v}^+) \leq C + C_0\rho(H(\bar{v})).$$

Notice first that this inequality follows from the explicit formulae in [12, Chapter II, Theorem 6.1] if the group \mathbb{G} has real rank one. One can also take $C_0 = 1$ in this case. In the general case, let X_1, X_2, \dots, X_r be a basis of the Lie algebra \mathfrak{n} with the following properties:

- (i) for any $1 \leq j \leq r$, the vector X_j belongs to a root space $\mathfrak{g}_{-\alpha}$ for some root $\alpha \in \Sigma^+$;
- (ii) if $\bar{\mathfrak{n}}_j$ denotes the linear span of the vectors X_1, \dots, X_j , then $\bar{\mathfrak{n}}_j$ is an ideal of \mathfrak{n} for any $1 \leq j \leq r$.

The map $\bar{v} : \mathbb{R}^r \rightarrow \overline{\mathfrak{N}}$,

$$\bar{v}(x_1, \dots, x_r) = (\exp x_r X_r) \cdots (\exp x_1 X_1)$$

gives a parametrization of the group $\overline{\mathfrak{N}}$. For any $\bar{v} = \bar{v}(x_1, \dots, x_r) \in \overline{\mathfrak{N}}$ let

$$\|\bar{v}\| = \max_j |x_j|.$$

A similar “norm” function on $\overline{\mathbb{N}}$ was used in [16]. It follows from [16, Lemma 5.3] that there exists a constant $c_1 > 0$ such that

$$e^{-2\rho(H(\overline{v}))} \leq C(2 + \|\overline{v}\|)^{-c_1}.$$

Thus

$$(3.10) \quad \log(2 + \|\overline{v}\|) \leq C + C_1\rho(H(\overline{v})).$$

We will now show that

$$(3.11) \quad |\overline{v}^+| \leq C \log(2 + \|\overline{v}\|).$$

Let $d(\cdot, \cdot)$ denote the distance function on the symmetric space \mathbb{X} induced by the Killing form on \mathfrak{g} . One has $|\overline{v}^+| = d(\overline{v}\mathbb{K}, e\mathbb{K})$. Therefore, if $\overline{v} = \overline{v}(x_1, \dots, x_r)$, then

$$|\overline{v}^+| \leq \sum_{j=1}^r d((\exp x_j X_j)\mathbb{K}, e\mathbb{K}).$$

It follows from the rank-one reduction method and the explicit rank-one formulae in [12, Chapter II, Theorem 6.1] that

$$d((\exp x_j X_j)\mathbb{K}, e\mathbb{K}) \leq C \log(2 + |x_j|).$$

Thus (3.11) follows from these last two estimates and (3.9) follows from (3.10), (3.11), and the observation that $\rho(\overline{v}^+) \leq |\rho||\overline{v}^+|$. \square

We will use this lemma to estimate the function $(\overline{v}, H) \rightarrow K_{p,\phi}(\overline{v}(\exp H))$ for any $\overline{v} \in \overline{\mathbb{N}}$ and $H \in \overline{\mathfrak{a}^+}$. Let $P(\overline{v}) = e^{-\rho(H(\overline{v}))}$; it is well known (see [10, Lemma 45]) that for any $\varepsilon > 0$,

$$(3.12) \quad \int_{\overline{\mathbb{N}}} P(\overline{v})^{(1+\varepsilon)} d\overline{v} = C_\varepsilon < \infty.$$

Lemma 3.4. (i) If $H \in \overline{\mathfrak{a}^+}$ and $\overline{v} \in \overline{\mathbb{N}}$, then

$$(3.13) \quad |K_{p,\phi}(\overline{v}(\exp H))| \leq \tilde{A}e^{-2\rho(H)/p}(1 + |H|)^{-(\ell-1+c_0)}P(\overline{v})^{2/p}.$$

(ii) If $H \in \overline{\mathfrak{a}^+}$ and $\overline{v} \in \overline{\mathbb{N}}$, then

$$(3.14) \quad K_{p,\phi}(\overline{v}(\exp H)) = e^{-2\rho(H)/p}\phi(H)P(\overline{v})^{2/p} + E_{p,\phi}(\overline{v}, H)$$

where

$$(3.15) \quad |E_{p,\phi}(\overline{v}, H)| \leq C \cdot \tilde{A}e^{-2\rho(H)/p}(1 + |H|)^{-(\ell-1+c_0)}P(\overline{v})^{2/p} \\ \times \left(\frac{1 + \rho(H(\overline{v}))}{1 + |H|} + \min(1, e^{-2\omega(H)}e^{C_0\rho(H(\overline{v}))}) \right).$$

The constant C_0 is the same as in Lemma 3.3.

Proof of Lemma 3.4. Notice first that

$$\lim_{|H| \rightarrow \infty} \phi(H) = 0.$$

This follows immediately from (3.1) and the fact that $\|\cdot\|_{L^p(\mathfrak{a})}$ is finite. Thus it follows from (3.1) that

$$(3.16) \quad |\phi(H)| \leq \tilde{A}(1 + |H|)^{-(\ell-1+c_0)}$$

for any $H \in \overline{\mathfrak{a}^+}$. To prove part (i) of the proposition notice that

$$K_{p,\phi}(\overline{v}(\exp H)) = e^{-2\rho(H+H(\overline{v})+H'(\overline{v},H))/p}\phi(H+H(\overline{v})+H'(\overline{v},H)).$$

Recall that $H(\overline{v})$ and $H'(\overline{v},H)$ belong to $\overline{\mathfrak{a}^+}$ ([10, Lemma 43]). Also, it is well known that $\rho(\overline{\mathfrak{a}^+}) > 0$, i.e., $H_\rho \in \mathfrak{a}^+$. Thus (3.13) follows easily from (3.16).

To prove part (ii) notice that it follows from Lemma 3.3 that the difference between $K_{p,\phi}(\overline{v}(\exp H))$ and $e^{-2\rho(H)/p}\phi(H)P(\overline{v})^{2/p}$ is equal to

$$e^{-2\rho(H)/p}P(\overline{v})^{2/p}\left[e^{-2\rho(H'(\overline{v},H))/p}\phi(H+H(\overline{v})+H'(\overline{v},H))-\phi(H)\right].$$

Since $H(\overline{v})$ and $H'(\overline{v},H)$ belong to $\overline{\mathfrak{a}^+}$, one has

$$|H(\overline{v})+H'(\overline{v},H)|\leq C\rho(H(\overline{v}))+\rho(H'(\overline{v},H)).$$

One can now combine the estimates (3.8) and (3.16) and the basic inequality (3.1) to show that

$$\begin{aligned} |E_{p,\phi}(\overline{v},H)| &\leq C\cdot \widetilde{A}e^{-2\rho(H)/p}(1+|H|)^{-(\ell+c_0)}P(\overline{v})^{2/p} \\ &\quad \times \left[1+\rho(H(\overline{v}))+e^{-2\omega(H)}(1+|H|)e^{C_0\rho(H(\overline{v}))}\right]. \end{aligned}$$

In addition, it follows from (3.13) and (3.16) that

$$|E_{p,\phi}(\overline{v},H)|\leq \widetilde{A}e^{-2\rho(H)/p}(1+|H|)^{-(\ell-1+c_0)}P(\overline{v})^{2/p}$$

and the estimate (3.15) follows from the last two inequalities. □

We return to the proof of Theorem 3.1. Let χ be the characteristic function of the set $\overline{\mathfrak{a}^+}$; we decompose the integral $I_{p,\phi}(f,g)$ in (3.6) as $I_{p,\phi}(f,g)=I_{p,\phi}^1(f,g)+I_{p,\phi}^2(f,g)+I_{p,\phi}^3(f,g)$, where

$$\begin{aligned} I_{p,\phi}^1(f,g) &= \int_{\overline{\mathbb{N}}}\int_{\overline{\mathbb{N}}}\int_{\mathfrak{a}}\int_{\mathfrak{a}}f(\overline{m},H_1)K_{p,\phi}(\delta_{-H_1}(\overline{m}^{-1}\overline{n})(\exp(H_2-H_1))) \\ &\quad \times g(\overline{n},H_2)e^{2\rho(H_1+H_2)}(1-\chi(H_2-H_1))dH_1dH_2d\overline{m}d\overline{n}, \\ I_{p,\phi}^2(f,g) &= \int_{\overline{\mathbb{N}}}\int_{\overline{\mathbb{N}}}\int_{\mathfrak{a}}\int_{\mathfrak{a}}f(\overline{m},H_1)E_{p,\phi}(\delta_{-H_1}(\overline{m}^{-1}\overline{n}),(H_2-H_1)) \\ &\quad \times g(\overline{n},H_2)e^{2\rho(H_1+H_2)}\chi(H_2-H_1)dH_1dH_2d\overline{m}d\overline{n}, \end{aligned}$$

and

$$\begin{aligned} I_{p,\phi}^3(f,g) &= \int_{\overline{\mathbb{N}}}\int_{\overline{\mathbb{N}}}\int_{\mathfrak{a}}\int_{\mathfrak{a}}f(\overline{m},H_1)P(\delta_{-H_1}(\overline{m}^{-1}\overline{n}))e^{-2\rho(H_2-H_1)/p}\phi(H_2-H_1) \\ &\quad \times g(\overline{n},H_2)e^{2\rho(H_1+H_2)}\chi(H_2-H_1)dH_1dH_2d\overline{m}d\overline{n}. \end{aligned}$$

We record an elementary fact that will be used several times in the rest of the paper.

Lemma 3.5. *One has*

$$(3.18) \qquad \int_{\overline{\mathfrak{a}^+}}(1+|H|)^{-(\ell-1+\varepsilon_0)}e^{-\varepsilon_1\omega(H)}dH\leq C_{\varepsilon_0,\varepsilon_1}$$

for any $\varepsilon_0,\varepsilon_1>0$.

Lemma 3.6. *One has*

$$(3.19) \qquad |I_{p,\phi}^1(f,g)|\leq C_p\cdot \widetilde{A}\|f\|_{L^p(\overline{\mathbb{N}}\times\mathfrak{a},d\mu)}\|g\|_{L^{p'}(\overline{\mathbb{N}}\times\mathfrak{a},d\mu)}.$$

Proof of Lemma 3.6. Let

$$\begin{cases} F_1(H_1) = [\int_{\bar{\mathbb{N}}} |f(\bar{m}, H_1)|^p d\bar{m}]^{1/p}; \\ G_1(H_2) = [\int_{\bar{\mathbb{N}}} |g(\bar{n}, H_2)|^{p'} d\bar{n}]^{1/p'}. \end{cases}$$

By Hölder's inequality, the absolute value of $I_{p,\phi}^1(f, g)$ is dominated by

$$(3.20) \quad \int_{\mathfrak{a}} \int_{\mathfrak{a}} F_1(H_1) G_1(H_2) e^{2\rho(H_1+H_2)} (1 - \chi(H_2 - H_1)) \\ \times \left(\int_{\bar{\mathbb{N}}} |K_{p,\phi}(\delta_{-H_1}(\bar{v})(\exp(H_2 - H_1)))| d\bar{v} \right) dH_1 dH_2.$$

Recall that the map $\bar{n}_1 \rightarrow \bar{n}_2 = \delta_H(\bar{n}_1)$ is a dilation of $\bar{\mathbb{N}}$ with $d\bar{n}_2 = e^{-2\rho(H)} d\bar{n}_1$. In addition, it is well known that the Abel transform

$$\mathcal{A}_2 k(H) = e^{\rho(H)} \int_{\bar{\mathbb{N}}} k(\bar{v}(\exp H)) d\bar{v}$$

takes suitable \mathbb{K} -bi-invariant functions k to W -invariant functions on \mathfrak{a} . For any regular element $H \in \mathfrak{a}$ let $H^+ \in \mathfrak{a}^+$ be its representative in the positive Weyl chamber, i.e., $H^+ \in (W \cdot H) \cap \mathfrak{a}^+$. Thus, if $H \in \mathfrak{a}$ is regular, one has

$$\begin{aligned} \int_{\bar{\mathbb{N}}} |K_{p,\phi}(\bar{v}(\exp H))| d\bar{v} &= e^{\rho(H^+ - H)} \int_{\bar{\mathbb{N}}} |K_{p,\phi}(\bar{v}(\exp H^+))| d\bar{v} \\ &\leq C_p \cdot \tilde{A} e^{\rho(H^+)} e^{-\rho(H)} e^{-2\rho(H^+)/p} (1 + |H|)^{-(\ell-1+c_0)}. \end{aligned}$$

The last estimate is a consequence of (3.12) and (3.13). It follows after several simplifications that the absolute value of the integral in (3.20) is dominated by

$$\begin{aligned} C_p \cdot \tilde{A} \int_{\mathfrak{a}} \int_{\mathfrak{a}} F_1(H_1) G_1(H_2) (1 - \chi(H_2 - H_1)) (1 + |H_2 - H_1|)^{-(\ell-1+c_0)} \\ \times e^{-\rho((H_2-H_1)^+)(2/p-1)} e^{\rho(H_1+H_2)} dH_1 dH_2. \end{aligned}$$

The change of variable $H_2 = H_1 + H$ shows that this integral is equal to

$$\begin{aligned} C_p \cdot \tilde{A} \int_{\mathfrak{a}} \left(\int_{\mathfrak{a}} (F_1(H_1) e^{2\rho(H_1)/p}) (G_1(H_1 + H) e^{2\rho(H_1+H)/p'}) dH_1 \right) \\ \times (1 - \chi(H)) (1 + |H|)^{-(\ell-1+c_0)} e^{-\rho(H^+ - H)(2/p-1)} dH. \end{aligned}$$

By Hölder's inequality, the inner integral is dominated by

$$\left(\int_{\mathfrak{a}} |F_1(H_1)|^p e^{2\rho(H_1)} dH_1 \right)^{1/p} \left(\int_{\mathfrak{a}} |G_1(H_2)|^{p'} e^{2\rho(H_2)} dH_2 \right)^{1/p'},$$

which is equal to

$$\|f\|_{L^p(\bar{\mathbb{N}} \times \mathfrak{a}, d\mu)} \|g\|_{L^{p'}(\bar{\mathbb{N}} \times \mathfrak{a}, d\mu)}.$$

Thus one only needs to estimate

$$(3.21) \quad \int_{\mathfrak{a}} (1 - \chi(H)) (1 + |H|)^{-(\ell-1+c_0)} e^{-\rho(H^+ - H)(2/p-1)} dH.$$

Notice that if $H \notin \mathfrak{a}^+$, then $\rho(H^+ - H) \geq \omega(H)$, where $\omega(H)$ has the same meaning as in Lemma 3.3. Since $p < 2$, it follows from Lemma 3.5 that the integral in (3.21) is dominated by $C_p \cdot \tilde{A}$ and the lemma follows. \square

Lemma 3.7. *One has*

$$(3.22) \qquad |I^2_{p,\phi}(f,g)| \leq C_p \cdot \widetilde{A} \|f\|_{L^p(\overline{\mathbb{N}} \times \mathfrak{a}, d\mu)} \|g\|_{L^{p'}(\overline{\mathbb{N}} \times \mathfrak{a}, d\mu)}.$$

Proof of Lemma 3.7. By Hölder’s inequality, the absolute value of the integral $I^2_{p,\phi}(f,g)$ is dominated by

$$(3.23) \qquad \int_{\mathfrak{a}} \int_{\mathfrak{a}} F_1(H_1) G_1(H_2) e^{2\rho(H_1+H_2)} \chi(H_2 - H_1) \\ \times \left(\int_{\overline{\mathbb{N}}} |E_{p,\phi}(\delta_{-H_1}(\overline{v}), (H_2 - H_1))| d\overline{v} \right) dH_1 dH_2.$$

Notice first that

$$\int_{\overline{\mathbb{N}}} |E_{p,\phi}(\delta_{-H_1}(\overline{v}), (H_2 - H_1))| d\overline{v} = e^{-2\rho(H_1)} \int_{\overline{\mathbb{N}}} |E_{p,\phi}(\overline{v}, (H_2 - H_1))| d\overline{v}.$$

One can use (3.15) to estimate the integral of $E_{p,\phi}$. It follows from (3.12) that

$$\int_{\overline{\mathbb{N}}} P(\overline{v})^{2/p} (1 + \rho(H(\overline{v}))) d\overline{v} \leq C_p.$$

Also

$$\begin{aligned} \int_{\overline{\mathbb{N}}} P(\overline{v})^{2/p} \min(1, e^{-2\omega(H)} e^{C_0 \rho(H(\overline{v}))}) d\overline{v} \\ \leq e^{-\omega(H)} \int_{\rho(H(\overline{v})) \leq \omega(H)/C_0} P(\overline{v})^{2/p} d\overline{v} + \int_{\rho(H(\overline{v})) \geq \omega(H)/C_0} P(\overline{v})^{2/p} d\overline{v} \\ \leq C_p e^{-\omega(H)} + \int_{\overline{\mathbb{N}}} e^{-(1/2+1/p)\rho(H(\overline{v}))} e^{-(1/p-1/2)\omega(H)/C_0} d\overline{v} \\ \leq C_p e^{-c_p \omega(H)} \end{aligned}$$

where $c_p = (1/p - 1/2)/C_0 > 0$. It follows from the last two estimates and (3.15) that if $H \in \overline{\mathfrak{a}^+}$, then

$$(3.24) \qquad \int_{\overline{\mathbb{N}}} |E_{p,\phi}(\overline{v}, H)| d\overline{v} \leq C_p \cdot \widetilde{A} e^{-2\rho(H)/p} \Psi(H)$$

where

$$\Psi(H) = (1 + |H|)^{-(\ell+c_0)} + (1 + |H|)^{-(\ell-1+c_0)} e^{-c_p \omega(H)}.$$

Let $\Psi(H) = 0$ if $H \notin \overline{\mathfrak{a}^+}$. It follows from (3.18) that $\Psi \in L^1(\mathfrak{a})$, i.e.,

$$(3.25) \qquad \|\Psi\|_{L^1(\mathfrak{a})} \leq C_p \cdot \widetilde{A}.$$

One substitutes the estimate (3.24) in (3.23). The change of variable $H_2 = H_1 + H$ shows that the absolute value of $I^2_{p,\phi}(f,g)$ is dominated by

$$C_p \cdot \widetilde{A} \int_{\mathfrak{a}} \left(\int_{\mathfrak{a}} (F_1(H_1) e^{2\rho(H_1)/p}) (G_1(H_1 + H) e^{2\rho(H_1+H)/p'}) dH_1 \right) \chi(H) \Psi(H) dH.$$

The lemma now follows from this last estimate, Hölder’s inequality and (3.25). \square

Lemma 3.8. *One has*

$$(3.26) \qquad |I^3_{p,\phi}(f,g)| \leq C_p \| |\ast \phi| \|_{L^p(\mathfrak{a})} \|f\|_{L^p(\overline{\mathbb{N}} \times \mathfrak{a}, d\mu)} \|g\|_{L^{p'}(\overline{\mathbb{N}} \times \mathfrak{a}, d\mu)}.$$

Proof of Lemma 3.8. The change of variable $H_2 = H_1 + H$ shows that $I_{p,\phi}^3(f, g)$ is equal to

$$(3.27) \quad \int_{\mathbb{N}} \int_{\mathbb{N}} \int_{\mathfrak{a}} f(\overline{m}, H_1) G_2(\overline{n}, H_1) P(\delta_{-H_1}(\overline{m}^{-1}\overline{n}))^{2/p} e^{2\rho(H_1)} e^{2\rho(H_1)/p} dH_1 d\overline{m} d\overline{n}$$

where

$$G_2(\overline{n}, H_1) = \int_{\mathfrak{a}} g(\overline{n}, H_1 + H) e^{2\rho(H_1+H)/p'} \phi(H) dH.$$

One has

$$(3.28) \quad \left(\int_{\mathfrak{a}} |G_2(\overline{n}, H_1)|^{p'} dH_1 \right)^{1/p'} \leq ||| * \phi |||_{L^p(\mathfrak{a})} \left(\int_{\mathfrak{a}} |g(\overline{n}, H_2)|^{p'} e^{2\rho(H_2)} dH_2 \right)^{1/p'}.$$

Let

$$G_3(H_1) = \left(\int_{\mathbb{N}} |G_2(\overline{n}, H_1)|^{p'} d\overline{n} \right)^{1/p'}.$$

It follows by Hölder's inequality and (3.12) that the absolute value of the integral (3.27) is dominated by

$$\int_{\mathfrak{a}} F_1(H_1) G_3(H_1) e^{2\rho(H_1)/p} dH_1,$$

which, again by Hölder's inequality, is dominated by

$$\left(\int_{\mathfrak{a}} G_3(H_1)^{p'} dH_1 \right)^{1/p'} \left(\int_{\mathfrak{a}} F_1(H_1)^p e^{2\rho(H_1)} dH_1 \right)^{1/p}.$$

The estimate (3.26) follows. \square

The main estimate (3.5) is a consequence of (3.19), (3.22) and (3.26).

4. L^p FOURIER MULTIPLIERS

Recall that the Fourier transform on the symmetric space \mathbb{X} associates to any smooth compactly supported function f on \mathbb{X} a function $\tilde{f} : \mathfrak{a}_{\mathbb{C}}^* \times \mathbb{K}/\mathbb{M} \rightarrow \mathbb{C}$. By the Plancherel theorem, any bounded W -invariant multiplier $m : \mathfrak{a}^* \rightarrow \mathbb{C}$ defines a bounded operator T_m on $L^2(\mathbb{X})$ given by $\widetilde{T_m f}(\lambda, b) = m(\lambda) \tilde{f}(\lambda, b)$. Assume that $p \in (1, 2) \cup (2, \infty)$ and let $\rho_p = |2/p - 1|\rho$. Let

$$\mathcal{T}_p = \mathfrak{a}^* + i\text{co}(W \cdot \rho_p),$$

where $\text{co}(W \cdot \rho_p)$ denotes the interior of the convex hull of the set of points $\{w \cdot \rho_p : w \in W\}$.

Assume that one has a multiplier $m : \mathfrak{a}^* \rightarrow \mathbb{C}$ that extends to a bounded W -invariant holomorphic function in the interior of the tube \mathcal{T}_p . Assume, in addition, that m satisfies the differential inequalities

$$(4.1) \quad \left| \partial_{\lambda}^j m(\lambda) \right| \leq A_j d(\lambda)^{-j}$$

for any $j = 0, 1, \dots$ and $\lambda \in \mathcal{T}_p$. Here $d(\lambda)$ denotes the distance between the point $\lambda \in \mathcal{T}_p$ and the set $i(\mathfrak{a}^* \setminus \text{co}(W \cdot \rho_p))$. More precisely, if $\lambda = \eta + i\xi$, then $d(\lambda) = (|\eta|^2 + d(\xi, \mathfrak{a}^* \setminus \text{co}(W \cdot \rho_p))^2)^{1/2}$. As before, $\partial_{\lambda}^j m(\lambda)$ denotes any partial derivative of m of order j .

Theorem 4.1. *If $p \in (1, 2) \cup (2, \infty)$ and m satisfies the differential inequalities (4.1), then the operator T_m extends to a bounded operator on $L^p(\mathbb{X})$.*

An operator T_m defined by a multiplier m that satisfies (4.1) can be thought of as a singular integral operator on the symmetric space \mathbb{X} . The rest of this section will be devoted to proving Theorem 4.1. The letter C will denote constants that may depend on the group \mathbb{G} and finitely many of the constants A_j in (4.1), and C_p will denote constants that may also depend on p .

One can assume that $p \in (1, 2)$. In order to insure the convergence of the integrals throughout this section we will also assume that the multiplier $m(\lambda)$ is premultiplied with a factor of the form $e^{-\delta\langle\lambda, \lambda\rangle}$, where $0 < \delta \leq 1$. Our estimates will be uniform in δ ; once one proves suitable uniform estimates, standard limiting arguments allow one to pass to the general theorem. The multiplier m is holomorphic in the interior of the tube \mathcal{T}_p ; therefore we can assume that for any $\xi \in \text{co}(W \cdot \rho_p)$ the function $\eta \rightarrow m(\eta + i\xi)$ is a Schwartz function on \mathfrak{a}^* .

Let K be the \mathbb{K} -bi-invariant kernel of the operator T_m . By the inversion formula one has

$$(4.2) \quad K(\exp H) = C \int_{\mathfrak{a}^*} m(\lambda) \varphi_\lambda(\exp H) |\mathbf{c}(\lambda)|^{-2} d\lambda$$

for any $H \in \overline{\mathfrak{a}^+}$. The spherical functions φ_λ are defined in (2.4) and \mathbf{c} is the Harish-Chandra function. The idea of the proof is the following. Using smooth cutoff functions we will break up this kernel into three parts. The first part is supported near the origin of the group \mathbb{G} (i.e., in the set $\{\mathbb{K}(\exp H)\mathbb{K} : |H| \leq 2\}$). The analysis of this local part of the kernel is contained in [1, Section 4]. The second part of the kernel is supported along the walls of the positive Weyl chamber, i.e., in the set $\{\mathbb{K}(\exp H)\mathbb{K} : H \in \overline{\mathfrak{a}^+} \text{ and } \omega(H) \leq 2\}$. One can use the Herz majorizing principle (see Proposition 2.1 and (2.7)) together with the main idea in [1, Section 2] to deal with this part of the kernel. Finally, the main part of the kernel is supported away from the walls of the positive Weyl chamber, i.e., in the set $\{\mathbb{K}(\exp H)\mathbb{K} : |H| \geq 1 \text{ and } \omega(H) \geq 1\}$. The Herz majorizing principle fails to prove boundedness on $L^p(\mathbb{X})$ of the operator defined by this part of the kernel if the multiplier m has a singularity at the point $i\rho_p$. We use Theorem 3.1 and the Harish-Chandra expansion of the spherical functions φ_λ to control the norm of the operator defined by this part of the kernel.

Let $\psi_1 : \overline{\mathfrak{a}^+} \rightarrow [0, 1]$ be a smooth cutoff function such that $\psi_1(H) = 0$ if $|H| \leq 1$ and $\psi_1(H) = 1$ if $|H| \geq 2$. Also, let $\psi_2 : \overline{\mathfrak{a}^+} \rightarrow [0, 1]$ be a smooth cutoff function such that $\psi_2(H) = 0$ if $\omega(H) \leq 1$ and $\psi_2(H) = 1$ if $\omega(H) \geq 2$. The kernel K in (4.2) can be written as $K = K_1 + K_2 + K_3$, where

$$K_1(\exp H) = (1 - \psi_1(H))K(\exp H),$$

$$K_2(\exp H) = \psi_1(H)(1 - \psi_2(H))K(\exp H),$$

and

$$K_3(\exp H) = \psi_1(H)\psi_2(H)K(\exp H)$$

for any $H \in \overline{\mathfrak{a}^+}$. The kernels K_1 , K_2 and K_3 are of course \mathbb{K} -bi-invariant kernels on \mathbb{G} . It follows easily from [1, Corollary 17] that the kernel K_1 defines a bounded operator on $L^p(\mathbb{X})$, i.e.,

$$(4.3) \quad ||| * K_1 |||_{L^p(\mathbb{X})} \leq C_p.$$

It remains to prove similar estimates on the norms of the operators defined by the kernels K_2 and K_3 .

Lemma 4.2. *One has*

$$||| * K_2 |||_{L^p(\mathbb{X})} \leq C_p.$$

Proof of Lemma 4.2. This proof is based on the main idea in Anker's paper [1]. By Proposition 2.1 it suffices to prove that

$$(4.4) \quad \int_{\mathbb{G}} |K_2(g)| e^{-2\rho(g^+)/p'} dg \leq C_p.$$

As in [1], for any $r \geq 0$ let

$$V_r = \{H \in \mathfrak{a} : (w \cdot \rho)(H) < |\rho|r \text{ for any } w \in W\}$$

and

$$U_r = \mathbb{K}(\exp V_r)\mathbb{K}.$$

We will prove that for any integer $r \geq 0$ one has

$$(4.5) \quad \int_{U_{r+1} \setminus U_r} |K_2(g)| dg \leq C_p e^{2|\rho|r/p'} (1+r)^{-3/2}.$$

Notice that $e^{-2\rho(g^+)/p'} \approx e^{-2|\rho|r/p'}$ if $g \in U_{r+1} \setminus U_r$. Thus (4.4) follows from (4.5) by summation over r .

We need the following estimate on the measure of the set $(U_{r+1} \setminus U_r) \cap (\text{supp } K_2)$:

$$(4.6) \quad |(U_{r+1} \setminus U_r) \cap \text{supp } K_2| \leq C(1+r)^{\ell-2} e^{2|\rho|r}.$$

Notice that the power of $(1+r)$ in (4.6) is $\ell-2$ rather than $\ell-1$ (compare with the estimate in [1, Lemma 6]). This is because of the "small" support of the kernel K_2 , and this gain is essential for the proof of the lemma. To prove (4.6) notice that for any $H \in (V_{r+1} \setminus V_r) \cap \overline{\mathfrak{a}^+}$ one has

$$\delta(H) \leq e^{2\rho(H)} \leq C e^{2|\rho|r},$$

where $\delta(H)$ is the density measure defined in (2.1). In addition, one can show easily that the measure in \mathfrak{a} of the set $(V_{r+1} \setminus V_r) \cap \overline{\mathfrak{a}^+} \cap (\text{supp } (1 - \psi_2))$ is dominated by $C(1+r)^{\ell-2}$. The estimate (4.6) follows.

By Hölder's inequality and (4.6), one has

$$\int_{U_{r+1} \setminus U_r} |K_2(g)| dg \leq C(1+r)^{(\ell-2)/2} e^{|\rho|r} ||| K_2 |||_{L^2(U_{r+1} \setminus U_r)}$$

for any integer $r \geq 0$. Thus it suffices to prove that

$$(4.7) \quad ||| K_2 |||_{L^2(U_{r+1} \setminus U_r)} \leq C_p e^{-|\rho_p|r} (1+r)^{-(\ell+1)/2}.$$

Let \mathcal{H} denote the Fourier transform on the symmetric space \mathbb{X} defined in (2.3). It is well known that \mathcal{H} extends to an isomorphism between $\mathbf{S}(\mathbb{G}/\mathbb{K})$ (the L^2 Schwartz space of \mathbb{K} -bi-invariant functions on \mathbb{G}) and $\mathbf{S}(\mathfrak{a}^*)^W$ (the subspace of W -invariant functions in the Schwartz space $\mathbf{S}(\mathfrak{a}^*)$). See [1, Section 1] for the definition of $\mathbf{S}(\mathbb{G}/\mathbb{K})$ and references. For any $f \in C_c^\infty(\mathbb{G}/\mathbb{K})$, let $\mathcal{A}(f)$ denote the Abel transform

$$\mathcal{A}(f)(H) = e^{\rho(H)} \int_{\mathbb{N}} f((\exp H)n) dn.$$

The Abel transform extends to an isomorphism between $\mathbf{S}(\mathbb{G}/\mathbb{K})$ and $\mathbf{S}(\mathfrak{a})^W$ (see [1, Proposition 3]). Finally, let \mathcal{F} denote the Euclidean Fourier transform

$$\mathcal{F}(g)(\lambda) = \int_{\mathfrak{a}} g(H) e^{-i\lambda(H)} dH,$$

which is an isomorphism between $\mathbf{S}(\mathfrak{a})^W$ and $\mathbf{S}(\mathfrak{a}^*)^W$. In addition, for any $f \in \mathbf{S}(\mathbb{G}/\mathbb{K})$ one has

$$\mathcal{H}(f) = \mathcal{F}(\mathcal{A}(f)).$$

Let $L = \mathcal{F}^{-1}(m) = \mathcal{A}(K)$. Assume that $r \geq 2$ in (4.7) (only simple modifications are needed if $r = 0$ or $r = 1$) and let $\Omega_r : \mathfrak{a} \rightarrow [0, 1]$ be the W -invariant smooth cutoff functions defined in [1, Section 2]. These functions are supported in the complement of V_{r-1} and are equal to 1 in the complement of V_r . The key observation in [1, Section 2] is that

$$K(\exp H) = \mathcal{A}^{-1}(L \cdot \Omega_r)(H)$$

for any H outside U_r . This is a simple consequence of a support property of the Abel transform (see [1, Proposition 4]). Thus, by the Plancherel theorem, one has

$$\begin{aligned} \|K_2\|_{L^2(U_{r+1} \setminus U_r)} &\leq \|K\|_{L^2(\mathbb{G} \setminus U_r)} \leq C \|\mathcal{A}^{-1}(L \cdot \Omega_r)\|_{L^2(\mathbb{G})} \\ &\leq C \|\mathcal{F}(L \cdot \Omega_r)\|_{L^2(\mathfrak{a}^*, |\mathbf{c}(\lambda)|^{-2} d\lambda)}. \end{aligned}$$

One also has the estimate

$$|\mathbf{c}(\lambda)|^{-2} \leq C(1 + |\lambda|^2)^{2b}$$

for any $\lambda \in \mathfrak{a}^*$, where b is a fixed positive integer. By the Euclidean Plancherel theorem

$$\begin{aligned} \|\mathcal{F}(L \cdot \Omega_r)\|_{L^2(\mathfrak{a}^*, |\mathbf{c}(\lambda)|^{-2} d\lambda)} &\leq C \|\mathcal{F}(L \cdot \Omega_r)(\lambda)(1 + |\lambda|^2)^b\|_{L^2(\mathfrak{a}^*)} \\ &\leq C \sum_{j=0}^b \|\Delta_{\mathfrak{a}}^j(L \cdot \Omega_r)\|_{L^2(\mathfrak{a})}, \end{aligned}$$

where $\Delta_{\mathfrak{a}}$ is the Laplace–Beltrami operator on \mathfrak{a} . Thus it suffices to prove that for any integer $j \in \{0, 1, \dots, b\}$ one has

$$(4.8) \quad \|\Delta_{\mathfrak{a}}^j(L \cdot \Omega_r)\|_{L^2(\mathfrak{a})} \leq C_p e^{-|\rho_p|r} (1+r)^{-(\ell+1)/2}.$$

Recall that L is the inverse Euclidean Fourier transform of the multiplier m ; therefore,

$$(4.9) \quad L(H) = C \int_{\mathfrak{a}^*} m(\lambda) e^{i\lambda(H)} d\lambda.$$

Assume that $H \in \overline{\mathfrak{a}^+}$ and $\rho(H) \geq 1$. Since the multiplier m is holomorphic, one can shift the integration in (4.9) to the space $i(1-\varepsilon)\rho_p + \mathfrak{a}^*$, where $\varepsilon\rho(H) = 1$. One has

$$L(H) = C_p e^{-\rho_p(H)} \int_{\mathfrak{a}^*} m(\lambda + i(1-\varepsilon)\rho_p) e^{i\lambda(H)} d\lambda.$$

We integrate by parts N times in λ and use (4.1). Notice also that if $H \in \overline{+\mathfrak{a}}$, then $\rho(H) \approx |H|$. Therefore one has

$$\begin{aligned} |L(H)| &\leq C_p e^{-\rho_p(H)} |H|^{-2N} \int_{\mathfrak{a}^*} |\Delta_{\mathfrak{a}^*}^N m(\lambda + i(1-\varepsilon)\rho_p)| d\lambda \\ (4.10) \quad &\leq C_p e^{-\rho_p(H)} |H|^{-2N} \int_{\mathfrak{a}^*} (\varepsilon + |\lambda|)^{-2N} d\lambda \\ &\leq C_p e^{-\rho_p(H)} \rho(H)^{-\ell} \end{aligned}$$

if $2N \geq \ell + 1$ and $H \in \overline{+\mathfrak{a}} \cap (\mathfrak{a} \setminus V_{r-1})$. The kernel L is W -invariant and the measure (in \mathfrak{a}) of the set $V_{r'} \setminus V_{r'-1}$ is dominated by $(1+r')^{\ell-1}$ for any $r' \geq 1$. The estimate (4.8) follows for $j = 0$. One can easily repeat the argument for any $j \leq b$ and Lemma 4.2 follows. \square

It remains to prove the following estimate:

Lemma 4.3. *One has*

$$||| * K_3 |||_{L^p(\mathbb{X})} \leq C_p.$$

Proof of Lemma 4.3. We will use the Harish-Chandra expansion of the spherical functions $\varphi_\lambda(H)$. Let

$$\mathfrak{a}_+^* = \{\lambda \in \mathfrak{a}^* : H_\lambda \in \mathfrak{a}^+\}$$

and let

$$B = \{\lambda \in \mathfrak{a}^* : \alpha(H_\lambda) > -\varepsilon_0 \text{ for any } \alpha \in \Sigma^+\}$$

be an open neighborhood of $\overline{\mathfrak{a}_+^*}$. The constant ε_0 will be chosen small and strictly positive. Let Q be the positive lattice generated by the simple roots $\alpha \in \Sigma^+$. If $H \in \mathfrak{a}^+$ and $\lambda \in \mathfrak{a}^*$, then one has the absolutely converging expansion

$$(4.11) \quad |\mathbf{c}(\lambda)|^{-2} \varphi_\lambda(H) = e^{-\rho(H)} \sum_{q \in 2Q} e^{-q(H)} \sum_{w \in W} \mathbf{c}(-w \cdot \lambda)^{-1} \Gamma_q(w \cdot \lambda) e^{i(w \cdot \lambda)(H)}.$$

The coefficient Γ_0 is equal to 1; the other coefficients Γ_q are rational functions in λ and extend to holomorphic functions in the tube $\mathfrak{a}^* + iB$ if the constant ε_0 is chosen small enough. Moreover, there exist constants C and d such that

$$(4.12) \quad |\Gamma_q(\lambda)| \leq C(1 + |q|)^d$$

for any $\lambda \in \mathfrak{a}^* + iB$ and $q \in Q$. We substitute the expansion (4.11) into the inversion formula (4.2) and notice that the integrand is W -invariant. One has

$$(4.13) \quad K_3(\exp H) = C \psi_1(H) \psi_2(H) e^{-\rho(H)} \sum_{q \in Q} e^{-q(H)} \int_{\mathfrak{a}^*} m(\lambda) \mathbf{c}(-\lambda)^{-1} \Gamma_q(\lambda) e^{i\lambda(H)} d\lambda.$$

The main term in the sum above corresponds to $q = 0$. We use the Herz majorizing principle to estimate the L^p norms of the operators induced by the error terms in (4.13). For any $q \in Q$ let

$$(4.14) \quad K_3^q(\exp H) = C \psi_1(H) \psi_2(H) e^{-\rho(H)} e^{-q(H)} \int_{\mathfrak{a}^*} m(\lambda) \mathbf{c}(-\lambda)^{-1} \Gamma_q(\lambda) e^{i\lambda(H)} d\lambda.$$

In general (i.e., if $\ell \geq 2$) the functions Γ_q do not satisfy favorable symbol-type estimates on \mathfrak{a}^* except for the uniform inequalities (4.12). However, one can use the fact that they are holomorphic functions in the tube $\mathfrak{a}^* + iB$ to estimate the absolute value of $K_3^q(\exp H)$. This idea was used in a recent work of Anker and Ji

[3]. See also [2] for similar estimates. Let Σ^{++} denote the set of indivisible roots $\alpha \in \Sigma^+$ and let

$$(4.15) \quad P(\lambda) = \prod_{\alpha \in \Sigma^{++}} (1 - i\langle \alpha, \lambda \rangle)^{k_\alpha},$$

where k_α are certain large integers. The function \mathbf{c} can be computed explicitly (see, for example, [2, Section 2]). This explicit formula shows that the function $\lambda \rightarrow \mathbf{c}(-\lambda)^{-1}$ is holomorphic in the tube $\mathfrak{a}^* + iB$ if ε_0 is small enough. In addition, there are positive constants k_α' such that

$$|\mathbf{c}(-\lambda)^{-1}| \leq C \prod_{\alpha \in \Sigma^{++}} |1 - i\langle \alpha, \lambda \rangle|^{k_\alpha'}$$

for any $\lambda \in \mathfrak{a}^* + iB$. Thus one can fix the constants k_α in (4.15) so that the function $\mathbf{c}(-\lambda)^{-1}/P(\lambda)$ belongs to the space $H^2(T_B)$ —see [19, Chapter III] for the theory of H^p spaces on tubes. By (4.12) the function $\lambda \rightarrow \Gamma_q(\lambda)\mathbf{c}(-\lambda)^{-1}/P(\lambda)$ belongs to the space $H^2(T_B)$ and

$$\int_{\mathfrak{a}^*} |\Gamma_q(\eta + i\xi)\mathbf{c}(-\eta - i\xi)^{-1}/P(\eta + i\xi)|^2 d\eta \leq C(1 + |q|)^{2d}$$

for any $\xi \in B$. One has

$$\begin{aligned} K_3^q(\exp H) &= C\psi_1(H)\psi_2(H)e^{-\rho(H)}e^{-q(H)}\mathcal{F}^{-1}(\lambda \rightarrow m(\lambda)\mathbf{c}(-\lambda)^{-1}\Gamma_q(\lambda)) \\ &= C\psi(H)e^{-\rho(H)}e^{-q(H)}\mathcal{F}^{-1}(m \cdot P) * \mathcal{F}^{-1}(f_q)(H), \end{aligned}$$

where $\psi = \psi_1\psi_2$ and $f_q(\lambda) = \Gamma_q(\lambda)\mathbf{c}(-\lambda)^{-1}/P(\lambda)$. By [19, Theorem 3.1, Section III], the function $\mathcal{F}^{-1}(f_q)$ is supported in the cone $-\overline{+\mathfrak{a}}$ and satisfies

$$\int_{-\overline{+\mathfrak{a}}} |\mathcal{F}^{-1}(f_q)(H)|^2 dH \leq C(1 + |q|)^{2d}.$$

On the other hand, $\mathcal{F}^{-1}(m \cdot P)$ is a certain derivative (in H) of the function $L = \mathcal{F}^{-1}(m)$. An argument similar to the one at the end of Lemma 4.2 (see (4.10)) shows that

$$|\mathcal{F}^{-1}(m \cdot P)(H)| \leq C_p e^{-\rho_p(H)}(1 + \rho(H))^{-\ell}$$

for any $H \in \overline{+\mathfrak{a}}$ with the property that $|H| \geq 1$. Combining the last three equations and Hölder's inequality one has

$$\begin{aligned} (4.16) \quad & |K_3^q(\exp H)| \\ & \leq C_p \psi(H) e^{-\rho(H)} e^{-q(H)} \int_{\mathfrak{a}} |\mathcal{F}^{-1}(m \cdot P)(H - H')| |\mathcal{F}^{-1}(f_q)(H')| dH' \\ & \leq C_p (1 + |q|)^d \psi(H) e^{-\rho(H)} e^{-q(H)} \left(\int_{\overline{+\mathfrak{a}}} |\mathcal{F}^{-1}(m \cdot P)(H + H')|^2 dH' \right)^{1/2} \\ & \leq C_p (1 + |q|)^d \psi(H) e^{-2\rho(H)/p} e^{-|q|\omega(H)} (1 + \rho(H))^{-\ell} \end{aligned}$$

for any $H \in \overline{\mathfrak{a}^+}$. The Herz majorizing principle applies if $|q| \geq 1$. Recall that the support of the cutoff function ψ is included in the set $\{H \in \overline{\mathfrak{a}^+} : \omega(H) \geq 1\}$. By

(2.7) and Lemma 3.5 one has

$$\begin{aligned} ||| * K_3^q |||_{L^p(\mathbb{X})} &\leq C_p \int_{\mathfrak{a}^+} |K_3^q(\exp H)| \delta(H) e^{-2\rho(H)/p'} dH \\ &\leq C_p (1 + |q|)^d e^{-(|q|-1)} \int_{\mathfrak{a}^+} (1 + |H|)^{-\ell} e^{-\omega(H)} dH \\ &\leq C_p (1 + |q|)^d e^{-|q|}. \end{aligned}$$

It remains to deal with the kernel K_3^0 . We need one more reduction before we can apply the results in the previous section. For any $H \in \overline{\mathfrak{a}^+}$ let

$$\widetilde{K}_3^0(\exp H) = C\psi_1(H) e^{-\rho(H)} \int_{\mathfrak{a}^*} m(\lambda) \mathbf{c}(-\lambda)^{-1} e^{i\lambda(H)} d\lambda$$

(compare with (4.14)) and extend \widetilde{K}_3^0 to a \mathbb{K} -bi-invariant function on \mathbb{G} . Notice that the function $H \rightarrow \left(K_3^0(\exp H) - \widetilde{K}_3^0(\exp H)\right)$ is supported in the set $\{H \in \overline{\mathfrak{a}^+} : \omega(H) \leq 2\}$ and satisfies an estimate similar to (4.16) for $q = 0$. By (2.7) and Lemma 3.5 one has

$$||| * (K_3^0 - \widetilde{K}_3^0) |||_{L^p(\mathbb{X})} \leq C_p.$$

It remains to prove a similar inequality for the kernel \widetilde{K}_3^0 . This will be a consequence of Corollary 3.2. For any $H \in \overline{\mathfrak{a}^+}$ let

$$(4.17) \quad \phi(H) = e^{2\rho(H)/p} K_3^0(\exp H) = C\psi_1(H) e^{\rho_p(H)} \int_{\mathfrak{a}^*} m(\lambda) \mathbf{c}(-\lambda)^{-1} e^{i\lambda(H)} d\lambda$$

and $\phi(H) = 0$ if $H \notin \overline{\mathfrak{a}^+}$. We have to prove that ϕ satisfies conditions (3.3) and (3.4) in the previous section. Notice that for any $H \in \overline{\mathfrak{a}^+}$ one has

$$\begin{aligned} \phi(H) &= C\psi_1(H) e^{\rho_p(H)} \mathcal{F}^{-1}(m \cdot P) * \mathcal{F}^{-1}(f_0)(H) \\ &= C\psi_1(H) \int_{\mathfrak{a}} \left(e^{\rho_p(H-H')} \mathcal{F}^{-1}(m \cdot P)(H - H') \right) \left(e^{\rho_p(H')} \mathcal{F}^{-1}(f_0)(H') \right) dH'. \end{aligned}$$

Recall also that $\mathcal{F}^{-1}(f_0)$ is an L^2 functions supported in the set $-\overline{^+\mathfrak{a}}$. Therefore the estimate (3.3) is an easy consequence of the following estimate:

$$(4.18) \quad \partial_H^j \left(H \rightarrow e^{\rho_p(H)} \mathcal{F}^{-1}(m \cdot P)(H) \right) \leq C_p (1 + |H|)^{-\ell-j}$$

for any $H \in \overline{^+\mathfrak{a}}$ with the property that $\rho(H) \geq 1$ and for $j = 0$ or $j = 1$. Let

$$\widetilde{\phi}(H) = e^{\rho_p(H)} \mathcal{F}^{-1}(m \cdot P)(H).$$

The proof of (4.18) for $j = 0$ is identical to the proof of (4.10) in Lemma 4.2. We will now prove (4.18) with $j = 1$ and $H = H_0 \in \overline{^+\mathfrak{a}}$ away from the origin. Let $\varepsilon = \rho(H_0)^{-1}$ and notice that

$$\widetilde{\phi}(H) = e^{\rho_p(H)/\rho(H_0)} \int_{\mathfrak{a}^*} m(\lambda + i(1 - \varepsilon)\rho_p) P(\lambda + i(1 - \varepsilon)\rho_p) e^{i\lambda(H)} d\lambda$$

by a shift of integration to the space $\mathfrak{a}^* + i(1 - \varepsilon)\rho_p$. Fix $H_1 \in \mathfrak{a}$ with $|H_1| = 1$. One has

$$\left| \frac{\partial \tilde{\phi}}{\partial H_1}(H_0) \right| \leq C \rho(H_0)^{-1} \left| \int_{\mathfrak{a}^*} m(\lambda + i(1 - \varepsilon)\rho_p) P(\lambda + i(1 - \varepsilon)\rho_p) e^{i\lambda(H_0)} d\lambda \right| \\ + C \left| \int_{\mathfrak{a}^*} m(\lambda + i(1 - \varepsilon)\rho_p) P(\lambda + i(1 - \varepsilon)\rho_p) \lambda(H_1) e^{i\lambda(H_0)} d\lambda \right|.$$

Let C_1 be the degree of the polynomial P . The estimate (4.18) for $j = 1$ follows by integration by parts as in (4.10). One only needs to notice that, as a consequence of (4.1),

$$|\Delta_{\mathfrak{a}^*}^N(\lambda \rightarrow m(\lambda + i(1 - \varepsilon)\rho_p) P(\lambda + i(1 - \varepsilon)\rho_p))| \leq C_N(\varepsilon + |\lambda|)^{-2N}(1 + |\lambda|)^{C_1}$$

and

$$|\Delta_{\mathfrak{a}^*}^N(\lambda \rightarrow m(\lambda + i(1 - \varepsilon)\rho_p) P(\lambda + i(1 - \varepsilon)\rho_p) \lambda(H_1))| \\ \leq C_N(\varepsilon + |\lambda|)^{1-2N}(1 + |\lambda|)^{C_1}$$

for any integer $N \geq 0$, where, as before, $\Delta_{\mathfrak{a}^*}$ denotes the Laplace–Beltrami operator on \mathfrak{a}^* .

The proof of the cancellation condition (3.4) is somewhat delicate, mainly because the support of the function ϕ is restricted to the cone $\overline{\mathfrak{a}^+}$. We proceed as in [14]. Let $\mu : \mathbb{R} \rightarrow [0, 1]$ be a smooth cutoff function supported in the interval $[1, \infty)$ and equal to 1 in the interval $[2, \infty)$. Let Σ^{+++} be the set of simple roots in Σ^+ and for any $H \in \mathfrak{a}$ let

$$\Psi(H) = \prod_{\alpha \in \Sigma^{+++}} \mu(\alpha(H)).$$

Notice that we can replace the cutoff function ψ_1 in (4.17) with Ψ . Indeed, let

$$\phi_1(H) = C \Psi(H) e^{\rho_p(H)} \int_{\mathfrak{a}^*} m(\lambda) \mathbf{c}(-\lambda)^{-1} e^{i\lambda(H)} d\lambda$$

and notice that the function $\phi - \phi_1$ is supported in the set $\{H \in \overline{\mathfrak{a}^+} : \omega(H) \leq 2\}$ and satisfies the estimate (3.3) for $j = 0$. Thus $\phi - \phi_1$ is an L^1 function. It remains to prove that the function ϕ_1 satisfies the cancellation condition (3.4). Recall that

$$|\phi_1(H)| \leq C(1 + |H|)^{-\ell}$$

for any $H \in \mathfrak{a}$. Thus the cancellation condition (3.4) is equivalent to

$$(4.19) \quad \left| \int_{\mathfrak{a}} \phi_1(H) \gamma(\varepsilon H) \right| \leq C$$

for any $\varepsilon \in (0, 1/2]$, where $\gamma : \mathfrak{a} \rightarrow [0, 1]$ is a smooth cutoff function supported in the ball $\{H \in \mathfrak{a} : |H| \leq 2\}$ and equal to 1 in the ball $\{H \in \mathfrak{a} : |H| \leq 1\}$. Let $T = \mathcal{F}^{-1}(\Psi)$ be a distribution on \mathfrak{a}^* . By the same argument as in [14, Section 4], it suffices to prove that

$$(4.20) \quad |T(\eta \rightarrow \tilde{m}(\eta - \eta_0 + i(1 - \varepsilon)\rho_p))| \leq C(1 + |\eta_0|)^{C_1}$$

for any $\eta_0 \in \mathfrak{a}^*$ and any $\varepsilon \in (0, 1]$, where C_1 is the same constant as before and $\tilde{m}(\lambda) = m(\lambda) \mathbf{c}(-\lambda)^{-1}$. For any $\alpha \in \Sigma^{+++}$ let $T_\alpha = \mathcal{F}^{-1}(H \rightarrow \mu(\alpha(H)))$. One has

$$T(f) = T_{\alpha_1} * \cdots * T_{\alpha_\ell}(f)$$

for any Schwartz function $f : \mathfrak{a} \rightarrow \mathbb{C}$, where $\alpha_1, \dots, \alpha_\ell$ are the positive simple roots. The estimate (4.20) will be a consequence of the following lemma:

Lemma 4.4. Assume that D_1 is an open set in the Euclidean space \mathbb{R}^n , $u \in \mathbb{R}^n$ is a fixed vector with $|u| = 1$ and D_2 is an open set with the property that $D_2 + tu \subset D_1$ for any $t \in [0, \delta_0]$, $\delta_0 > 0$. Let T_u be the inverse Fourier transform of the function $v \rightarrow \mu(\langle u, v \rangle)$ (in the sense of distributions). Assume that f is a holomorphic function in the tube $\mathbb{R}^n + iD_1$ with the property that

$$|f(x + iy)| \leq A(1 + |x|)^{\delta_1}$$

for any $x \in \mathbb{R}^n$ and $y \in D_1$. In addition, assume that the function $x \rightarrow f(x + iy)$ is a Schwartz function on \mathbb{R}^n for any $y \in D_1$. Then the function

$$f * T_u(x + iy) = T_u(v \rightarrow f(x - v + iy))$$

has the property that

$$|f * T_u(x + iy)| \leq C_{\delta_0} \cdot A(1 + |x|)^{\delta_1}$$

for any $x \in \mathbb{R}^n$ and $y \in D_2$. In addition, the function $x \rightarrow f * T_u(x + iy)$ is a Schwartz function on \mathbb{R}^n for any $y \in D_2$. The constant C_{δ_0} depends only on δ_0 and the cutoff function μ .

The estimate (4.20) follows from this lemma by a simple inductive argument. One starts with $f(\lambda) = \tilde{m}(-\lambda)$ and $D_1 = -\text{co}(W \cdot \rho_p)$ and applies Lemma 4.4 to the vectors $u = H_{\alpha_1}, \dots, H_{\alpha_\ell}$. One only needs to check that there exists a small constant $\delta_0 = \delta_0(p)$ with the property that

$$(4.21) \quad (1 - \varepsilon)\rho_p - (t_1\alpha_1 + \dots + t_\ell\alpha_\ell) \in \text{co}(W \cdot \rho_p)$$

for any $t_1, \dots, t_\ell \in [0, \delta_0]$. Recall that $\varepsilon \in (0, 1/2]$. We need the following well-known fact about convex hulls (see [11, Chapter IV]): if $H \in \overline{\mathfrak{a}^+}$ and $C(H)$ is the closure of the orbit $\{w \cdot H : w \in W\}$, then

$$C(H) \cap \overline{\mathfrak{a}^+} = (H - \overline{+\mathfrak{a}}) \cap \overline{\mathfrak{a}^+}.$$

We can pass this to the space \mathfrak{a}^* and apply it to the vector ρ . One has

$$C(\rho) \cap \overline{\mathfrak{a}_+^*} = (\rho - \overline{+\mathfrak{a}^*}) \cap \overline{\mathfrak{a}_+^*}$$

where $C(\rho)$ is the closure of the set $\text{co}(W \cdot \rho)$ and, as before, \mathfrak{a}_+^* and $+\mathfrak{a}^*$ are the cones corresponding to \mathfrak{a}^+ and $+\mathfrak{a}$, respectively. Notice also that $\alpha_1, \dots, \alpha_\ell \in \overline{+\mathfrak{a}^*}$. Thus one has

$$\rho - (t_1\alpha_1 + \dots + t_\ell\alpha_\ell) \in C(\rho)$$

for any $t_1, \dots, t_\ell \in [0, c]$ and (4.21) follows. \square

Proof of Lemma 4.4. This is a consequence of the one-dimensional estimate proved in [14, Section 4]. Assume that $u = (1, 0, \dots, 0)$. The distribution T_u is the inverse Fourier transform of the function $v \rightarrow \mu(v_1)$. The same argument as in [14, Section 4] shows that

$$T_u(f) = \int_{\mathbb{R}} \frac{\nu(t)(f(t, 0, \dots, 0) - f(0)e^{-t^2})}{-it} dt + Cf(0)$$

for any Schwartz function f , where ν is the inverse Fourier transform of the function μ' . Thus

$$(4.22) \quad f * T_u(x + iy) = \int_{\mathbb{R}} \frac{\nu(t)(f(x - tu + iy) - f(x + iy)e^{-t^2})}{-it} dt + Cf(x + iy)$$

where f is the function in Lemma 4.4 and $y \in D_2$. The function ν , the inverse Fourier transform of the smooth compactly supported function μ' , extends to a holomorphic function in the plane. We shift the integration in (4.22) to the line $-i\delta_0 + \mathbb{R}$, where δ_0 is the constant in Lemma 4.4. Notice also that the function $t \rightarrow \nu(t - i\delta_0)$ is a Schwartz function on \mathbb{R} and the lemma follows. \square

REFERENCES

- [1] J.-Ph. Anker, *L^p Fourier multipliers on Riemannian symmetric spaces of the noncompact type*, Ann. of Math. **132** (1990), 597–628. MR **92e**:43006
- [2] J.-Ph. Anker and L. Ji, *Heat kernel and Green function estimates on noncompact symmetric spaces*, Geom. Funct. Anal. **9** (1999), 1035–1091. MR **2001b**:58038
- [3] J.-Ph. Anker and L. Ji, *Heat kernel and Green function estimates on noncompact symmetric spaces II*, Preprint (1999).
- [4] J.-Ph. Anker and N. Lohoué, *Multiplicateurs sur certains espaces symétriques*, Amer. J. Math. **108** (1986), 1303–1354. MR **88c**:43008
- [5] J.-L. Clerc and E. M. Stein, *L^p -multipliers for noncompact symmetric spaces*, Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 3911–3912. MR **51**:3803
- [6] R. Coifman and G. Weiss, *Transference Methods in Analysis*, CBMS Regional Conference Series in Mathematics, No. 31, Amer. Math. Soc., Providence, RI (1976). MR **58**:2019
- [7] M. Cowling, *The Kunze–Stein phenomenon*, Ann. Math. **107** (1978), 209–234. MR **58**:22398
- [8] M. Cowling, S. Giulini and S. Meda, *$L^p - L^q$ estimates for functions of the Laplace–Beltrami operator on noncompact symmetric spaces. I*, Duke Math. J. **72** (1993), 109–150. MR **95b**:22031
- [9] S. Giulini, G. Mauceri and S. Meda, *L^p multipliers on noncompact symmetric spaces*, J. reine angew. Math. **482** (1997), 151–175. MR **98g**:43006
- [10] Harish-Chandra, *Spherical functions on a semisimple Lie group. I*, Amer. J. Math. **80** (1958), 241–310. MR **20**:925
- [11] S. Helgason, *Groups and Geometric Analysis; Integral Geometry, Invariant Differential Operators and Spherical Functions*, Academic Press, New York (1984). MR **86c**:22017
- [12] S. Helgason, *Geometric Analysis on Symmetric Spaces*, Amer. Math. Soc., Providence, RI (1994). MR **96h**:43009
- [13] C. Herz, *Sur le phénomène de Kunze–Stein*, C. R. Acad. Sci. Paris, Série A **271** (1970), 491–493. MR **43**:6741
- [14] A. D. Ionescu, *Singular integrals on symmetric spaces of real rank one*, Duke Math. J. **114** (2002), 101–122. MR **2003c**:43008
- [15] R. A. Kunze and E. M. Stein, *Uniformly bounded representations and harmonic analysis of the 2×2 unimodular group*, Amer. J. Math. **82** (1960), 1–62. MR **29**:1287
- [16] L.-A. Lindahl, *Fatou’s theorem for symmetric spaces*, Ark. Mat. **10** (1972), 33–47. MR **52**:3892
- [17] N. Lohoué and T. Rychener, *Some function spaces on symmetric spaces related to convolution operators*, J. Funct. Anal. **55** (1984), 200–219. MR **85d**:22024
- [18] R. J. Stanton and P. A. Tomas, *Expansions for spherical functions on noncompact symmetric spaces*, Acta Math. **140** (1978), 251–271. MR **58**:23365
- [19] E. M. Stein, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Univ. Press (1971). MR **46**:4102
- [20] E. M. Stein, *Harmonic Analysis*, Princeton Univ. Press (1993). MR **95c**:42002
- [21] J. O. Strömberg, *Weak type L^1 estimates for maximal functions on noncompact symmetric spaces*, Ann. Math. **114** (1981), 115–126. MR **82k**:43010
- [22] M. E. Taylor, *L^p estimates on functions of the Laplace operator*, Duke Math. J. **58** (1989), 773–793. MR **91d**:58253

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

E-mail address: aionescu@math.mit.edu

Current address: University of Wisconsin – Madison, Madison, Wisconsin 53706

E-mail address: ionescu@math.wisc.edu

WEST'S PROBLEM ON EQUIVARIANT HYPERSPACES AND BANACH-MAZUR COMPACTA

SERGEY ANTONYAN

ABSTRACT. Let G be a compact Lie group, X a metric G -space, and $\exp X$ the hyperspace of all nonempty compact subsets of X endowed with the Hausdorff metric topology and with the induced action of G . We prove that the following three assertions are equivalent: (a) X is locally continuum-connected (resp., connected and locally continuum-connected); (b) $\exp X$ is a G -ANR (resp., a G -AR); (c) $(\exp X)/G$ is an ANR (resp., an AR). This is applied to show that $(\exp G)/G$ is an ANR (resp., an AR) for each compact (resp., connected) Lie group G . If G is a finite group, then $(\exp X)/G$ is a Hilbert cube whenever X is a nondegenerate Peano continuum. Let $L(n)$ be the hyperspace of all centrally symmetric, compact, convex bodies $A \subset \mathbb{R}^n$, $n \geq 2$, for which the ordinary Euclidean unit ball is the ellipsoid of minimal volume containing A , and let $L_0(n)$ be the complement of the unique $O(n)$ -fixed point in $L(n)$. We prove that: (1) for each closed subgroup $H \subset O(n)$, $L_0(n)/H$ is a Hilbert cube manifold; (2) for each closed subgroup $K \subset O(n)$ acting non-transitively on S^{n-1} , the K -orbit space $L(n)/K$ and the K -fixed point set $L(n)[K]$ are Hilbert cubes. As an application we establish new topological models for the Banach-Mazur compacta $L(n)/O(n)$ and prove that $L_0(n)$ and $(\exp S^{n-1}) \setminus \{S^{n-1}\}$ have the same $O(n)$ -homotopy type.

1. INTRODUCTION

In 1976 J. E. West [33] asked the following question: Let G be a compact, connected Lie group. Is the orbit space $(\exp G)/G$ an absolute retract, and if so, is it always homeomorphic to the Hilbert cube? In a more general form this problem appeared also in [34, Problem 1022].

These questions have remained open except when $G = S^1$, the circle group, where the answers are “Yes” and “No”, respectively. Toruńczyk and West proved in [29] that the orbit space $(\exp_0 S^1)/S^1$ is an Eilenberg-MacLane space $\mathbf{K}(\mathbb{Q}, 2)$, where S^1 is the circle group and \mathbb{Q} stands for the rationals.

Recall that if G is a compact group and X a metrizable G -space, then $\exp X$ denotes the hyperspace of all non-void compact subsets of X , equipped with the Hausdorff metric topology and with the induced action of G . We use $\exp_0 X$ for the complement $(\exp X) \setminus \{X\}$.

Received by the editors May 1, 2000 and, in revised form, September, 15, 2002.

2000 *Mathematics Subject Classification.* Primary 57N20, 57S10, 54B20, 54C55, 55P91, 46B99.

Key words and phrases. Banach-Mazur compacta, G -ANR, Q -manifold, hyperspace, orbit space, homotopy type, G -nerve.

The author was supported in part by grant IN-105800 from PAPIIT (UNAM).

In the first part of the present paper we give a positive answer to the first question of West's problem as a corollary to the following general result.

Theorem 1.1. *Let G be a compact Lie group and X a metrizable G -space. Then the following are equivalent:*

- (1) X is locally continuum-connected (resp., connected and locally continuum-connected),
- (2) $\exp X$ is a G -ANR (resp., a G -AR),
- (3) The orbit space $(\exp X)/G$ is an ANR (resp., an AR).

The proof of this theorem is given in Section 3.

Since each compact Lie group G is locally path-connected, Theorem 1.1, in particular, yields a positive answer to the first question of West's Problem (Corollary 3.8).

Remark 1.2. As is shown in Section 3, the implications $(1) \implies (2) \implies (3)$ in Theorem 1.1 are true also for arbitrary compact (not necessarily Lie) groups; $(1) \implies (2)$ and $(1) \implies (3)$ can be regarded as equivariant versions of Wojdyslawski's Theorem [35] as sharpened by D. Curtis [13].

In Heisey and West [17] it was proved that if G is a finite group and X is a nondegenerate Peano continuum, then $(\exp X)/G$ is a Hilbert cube if it is an AR. Consequently, in combination with Theorem 1.1, this implies that $(\exp X)/G$ is always a Hilbert cube whenever X is a nondegenerate Peano continuum (Corollary 3.9).

The second part of this paper is devoted to the Banach-Mazur compacta.

In his 1932 book *Théorie des Opérations Linéaires*, S. Banach [10] introduced, for each $n \geq 2$, the space of isometry classes $[E]$ of n -dimensional Banach spaces equipped with the metric

$$d([E], [F]) = \ln \inf \{ \|T\| \cdot \|T^{-1}\| \mid T : E \rightarrow F \text{ is a linear isomorphism} \}.$$

These spaces are now denoted by $BM(n)$ and called the Banach-Mazur compacta. The topology of these spaces continues to be of interest, and the following questions of A. Pelczyński were included in J. West's list of problems in the 1990 book *Open Problems in Topology* [34, Problem 899]: (1) Are the Banach-Mazur compacta $BM(n)$ AR's? (2) Are they Hilbert cubes?

The AR part of Pelczyński's problem has been solved affirmatively due to efforts of P. Fabel [16] and the author [8]. While the question of whether the Banach-Mazur compacta $BM(n)$ are homeomorphic to the Hilbert cube remains open for all $n \geq 3$, it was answered negatively for $n = 2$ in [9].

In the second part of the paper we establish some new properties of the Banach-Mazur compacta and consider their relation to West's problem above. It turns out that Pelczyński's problem and West's problem are of the same nature, and both problems can be considered from a unified point of view.

We recall some necessary notation first. By $\mathcal{B}(n)$ we denote the hyperspace of all centrally symmetric (about the origin), compact, convex bodies in \mathbb{R}^n . We consider the Hausdorff metric topology on $\mathcal{B}(n)$ and the natural induced action of the full linear group $GL(n)$ on it. As usual, we shall use $O(n)$ for the orthogonal group. It is well known that $BM(n)$ is homeomorphic to the orbit space $\mathcal{B}(n)/GL(n)$ (see [34, p. 544]).

According to a classical theorem of F. John [21], for any $A \in \mathcal{B}(n)$ there is a unique minimal-volume ellipsoid $l(A)$ containing A (respectively, maximal-volume ellipsoid $j(A)$ contained in A). Usually $j(A)$ is called the John ellipsoid of A and $l(A)$ is called the Löwner ellipsoid of A .

Let $J(n)$ and $L(n)$ be the $O(n)$ -invariant subsets of $\mathcal{B}(n)$ consisting of all bodies $A \in \mathcal{B}(n)$ for which the ordinary Euclidean unit ball

$$B^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + \dots + x_n^2 \leq 1\}$$

is the John ellipsoid and the Löwner ellipsoid, respectively. Some properties of these $O(n)$ -spaces are studied in [9]. It was proved in [9, Theorem 4 and Remark 1] that $J(n)$ and $L(n)$ are global compact $O(n)$ -slices for the $GL(n)$ -space $\mathcal{B}(n)$. It then follows from a result of H. Abels [1, Lemma 2.3] that $J(n)$ and $L(n)$ are $O(n)$ -homeomorphic. The Banach-Mazur compactum $BM(n)$ is just the orbit space $J(n)/O(n)$ [9, Corollary 1] or, equivalently, the orbit space $L(n)/O(n)$ (see [9, Remark 1]). In what follows we shall use the model $BM(n) = L(n)/O(n)$. By $L_0(n)$ we denote the complement $L(n) \setminus \{B^n\}$, and $BM_0(n) = L_0(n)/O(n)$. As usual, we reserve the letter Q for the Hilbert cube.

Here we prove four basic properties about the Banach-Mazur compacta $BM(n)$, $n \geq 2$.

Theorem 1.3. *For any closed subgroup $H \subset O(n)$, the orbit space $L_0(n)/H$ is a $[0, 1)$ -stable Q -manifold. In particular, $BM_0(n)$ is a $[0, 1)$ -stable Q -manifold.*

Recall that a Q -manifold is said to be $[0, 1)$ -stable if it is homeomorphic to its product with the half-open interval $[0, 1)$ (see [12, Ch. V]).

Theorem 1.4. *For each closed subgroup $K \subset O(n)$ acting non-transitively on S^{n-1} , the K -orbit space $L(n)/K$, as well as the K -fixed point set $L(n)[K]$, is a Hilbert cube. In particular, $L(n)$ is a Hilbert cube.*

These theorems are proved in Section 5. However, Section 4 should also be considered as a part of those proofs, because the technique we develop there is further applied to Theorems 1.3 and 1.4.

Remark 1.5. Below, in Lemma 7.4, the hypothesis that K acts non-transitively on the sphere S^{n-1} is shown to be equivalent to the condition $L_0(n)[K] \neq \emptyset$.

Next, in Section 6 we apply Theorem 1.3 to establish a new topological model for the Banach-Mazur compacta $BM(n)$ for arbitrary $n \geq 2$.

Namely, assume that $(H_1), (H_2), \dots$ is the sequence of all $O(n)$ -orbit types occurring in $L_0(n)$. Let $\text{Cone}(O(n)/H_i)$ denote the cone over $O(n)/H_i$ endowed with the quotient topology and with the translation action of $O(n)$ on the levels. Let

$$\Pi(n) = \prod_{i=1}^{\infty} Q(H_i), \quad \text{where } Q(H_i) = (\text{Cone}(O(n)/H_i))^{\infty},$$

each equipped with the diagonal $O(n)$ -action.

Denote $\Pi_0(n) = \Pi(n) \setminus \{a\}$, where a is the unique $O(n)$ -fixed point of $\Pi(n)$. Since $\text{Cone}(O(n)/H_i) \in \text{AR}$, $i \geq 1$, it then follows from a result of West [32] that $\Pi(n)$ is a Hilbert cube.

Theorem 1.6. *For each closed subgroup $H \subset O(n)$, the two H -orbit spaces $L_0(n)/H$ and $\Pi_0(n)/H$ are homeomorphic. In particular, the Banach-Mazur compactum $BM(n)$ is homeomorphic to the orbit space $\Pi(n)/O(n)$.*

Remark 1.7. In combination with [9, Corollary 10], Theorem 1.6 gives yet simpler topological models for $BM(2)$.

The idea that Pelczyński's problem on Banach-Mazur compacta is closely related to West's problem on equivariant hyperspaces was expressed in [9] in the following form:

Conjecture 1.8. *For each closed subgroup $H \subset O(n)$, $n \geq 2$, the orbit spaces $L_0(n)/H$ and $(\exp_0 S^{n-1})/H$ are homeomorphic Q -manifolds. In particular, the Banach-Mazur compactum $BM(n)$ is homeomorphic to $(\exp S^{n-1})/O(n)$.*

Since the sphere S^{n-1} is $O(n)$ -homeomorphic to the coset space $O(n)/O(n-1)$, we see that $(\exp S^{n-1})/O(n)$ is just of the form $(\exp(G/H))/G$. So, if Conjecture 1.8 were proved, the Banach-Mazur compacta would be just of the form $(\exp(G/H))/G$ (with $G = O(n)$ and $H = O(n-1)$). On the other hand, the space $(\exp G)/G$ in West's problem is also of the form $(\exp(G/H))/G$ (with H the trivial subgroup). This shows how close are, in fact, Pelczyński's problem and West's problem.

Here is our fourth result on Banach-Mazur compacta, which is proved in Section 7:

Theorem 1.9. *$\exp_0 S^{n-1}$ and $L_0(n)$ have the same $O(n)$ -homotopy type.*

It follows immediately from Theorem 1.9 that $L_0(n)/H$ and $(\exp_0 S^{n-1})/H$ have the same homotopy type for any closed subgroup $H \subset O(n)$. This result and Theorem 1.3 constitute essential steps in proving Conjecture 1.8. However, we do not prove Conjecture 1.8 in this paper. The only step we lack to complete its proof is that $(\exp_0 S^{n-1})/H$, $n \geq 2$, is a Q -manifold. The details of this reduction are also presented in Section 6 (Theorem 7.9).

The paper is divided as follows:

- §1. Introduction.
- §2. Preliminaries.
- §3. Proof of Theorem 1.1 and its corollaries.
- §4. The G -nerve.
- §5. Proofs of Theorems 1.3 and 1.4.
- §6. Proof of Theorem 1.6.
- §7. Proof of Theorem 1.9 and reduction of Conjecture 1.8.

2. PRELIMINARIES

For a given topological group G , we denote by $G\text{-A(N)R}$ (resp., by $G\text{-A(N)E}$) the class of all G -equivariant absolute (neighborhood) retracts (resp., extensors) for all metrizable G -spaces. These concepts are straightforward extensions to the case of G -spaces of the corresponding concepts of ordinary $A(N)R$'s and $A(N)E$'s (see, for example, [2]–[6]). We refer to the monographs [11] and [26] for basic notions of the theory of G -spaces.

If G is a topological group and X is a G -space, for any $x \in X$ we denote the stabilizer (or stationary subgroup) of x by $G_x = \{g \in G \mid gx = x\}$.

For each subgroup $H \subset G$, the H -fixed point set $X[H]$ is defined to be the set $\{x \in X \mid H \subset G_x\}$.

The family of all subgroups of G that are conjugate to H is denoted by (H) , i.e., $(H) = \{gHg^{-1} \mid g \in G\}$. We will call (H) a G -orbit type (or simply an orbit type).

For two orbit types (H_1) and (H_2) , one says that $(H_1) \preceq (H_2)$ iff $H_1 \subset g^{-1}H_2g$ for some $g \in G$. The relation \preceq is a partial ordering on the set of all orbit types. Since $G_{gx} = gG_xg^{-1}$ for any $x \in X$ and $g \in G$, we have $(G_x) = \{G_{gx} \mid g \in G\}$.

For a subset $S \subset X$, $H(S)$ denotes the H -saturation of S , i.e., $H(S) = \{hs \mid h \in H, s \in S\}$. In particular, $H(x)$ denotes the H -orbit $\{hx \in X \mid h \in H\}$ of x . The H -orbit space is denoted by X/H . In particular, X/G denotes the orbit space of X .

By G/H we will denote the G -space of cosets $\{gH \mid g \in G\}$ under the action of G induced by left translations.

A continuous map $f : X \rightarrow Y$ of G -spaces is said to be equivariant or G -equivariant or, for short, a G -map, if $f(gx) = gf(x)$ for all $g \in G$, $x \in X$. An equivariant map $f : X \rightarrow Y$ is said to be *isovariant* (or G -isovariant) if $G_x = G_{f(x)}$ for all $x \in X$.

A compatible metric ρ on a G -space is called invariant or G -invariant if $\rho(gx, gy) = \rho(x, y)$ for all $x, y \in X$ and $g \in G$.

If X is metrized by a G -invariant metric ρ , then the formula $\tilde{\rho}(G(x), G(y)) = \inf\{\rho(x', y') \mid x' \in G(x), y' \in G(y)\}$ defines a metric $\tilde{\rho}$, compatible with the quotient topology of X/G , whenever G is a compact group.

Let us recall the well-known and important definition of a slice [26]:

Definition 2.1. Let G be a topological group, $H \subset G$ a closed subgroup and X a G -space. A subset $S \subset X$ is called an H -slice in X if

- (1) S is H -invariant, i.e., $H(S) = S$,
- (2) the saturation $G(S)$ is open in X ,
- (3) if $g \in G \setminus H$, then $gS \cap S = \emptyset$, and
- (4) S is closed in $G(S)$.

If in addition $G(S) = X$, then we say that S is a *global* H -slice of X .

The following is one of the fundamental results in topological transformation group theory (see [26, Corollary 1.7.19, Corollary 1.7.20 and Theorem 1.7.7] or [11, Ch. II, §§4 and 5]):

Theorem 2.2 (Slice theorem). *Let G be a compact Lie group, X a Tychonoff G -space and $x \in X$ any point. Then:*

- (1) *There exists a G_x -slice $S \subset X$ such that $x \in S$.*
- (2) *$(G_y) \preceq (G_x)$ for each point $y \in G(S)$.*
- (3) *There exists a unique G -map $f : G(S) \rightarrow G/G_x$ such that $S = f^{-1}(eG_x)$.*

In [9], using the classical result of John [21] on the minimal-volume ellipsoid, it was proved that $L(n)$ is a global $O(n)$ -slice for the $GL(n)$ -space $\mathcal{B}(n)$. In combination with a result of H. Abels [1, Theorem 2.1] this yields the following theorem, which we will need in the sequel:

Theorem 2.3. *There is an $O(n)$ -equivariant retraction $r : \mathcal{B}(n) \rightarrow L(n)$ such that $r(A)$ belongs to the $GL(n)$ -orbit $GL(n)(A)$ for every $A \in \mathcal{B}(n)$.*

In [9, Corollary 2] it was proved that $J(n)$ is a compact $O(n)$ -AR, and since $L(n)$ is $O(n)$ -homeomorphic to $J(n)$ [9, Remark 1], we have the following result that will often be used in what follows:

Theorem 2.4 ([9]). *$L(n)$ is a compact $O(n)$ -AR.*

Let $f_0, f_1: X \rightarrow X'$ be G -maps. A G -homotopy of f_0 into f_1 is a homotopy in the ordinary sense which is a G -map at each stage of the deformation. A G -space X is called G -contractible if there is a G -fixed point $* \in X$ such that the constant G -map $X \rightarrow \{*\}$ and the identity map 1_X are G -homotopic. A G -map $f: X \rightarrow Y$ is a G -homotopy equivalence if there is a G -map $f': Y \rightarrow X$ such that $f'f$ is G -homotopic to 1_X and ff' is G -homotopic to 1_Y .

Theorem 2.5 ([20]). *Let G be a compact Lie group and $f: T \rightarrow Z$ a G -map of G -ANR's. Then f is a G -homotopy equivalence iff for each closed subgroup $K \subset G$, the restriction of f to the K -fixed point set $T[K]$ is an ordinary homotopy equivalence.*

Remark 2.6. In [20, Proposition 4.1] the result originally was stated for paracompact G -ANE's (even in its fiberwise form). However, the proof in [20] serves for metrizable G -ANR's as well.

Yet another basic result for this paper is the following.

Theorem 2.7 ([6], [7]). *Let G be a compact group, $N \subset G$ a closed normal subgroup and X a G -ANR (resp., a G -AR). Then the N -orbit space X/N , endowed with the induced action of the quotient group G/N , is a G/N -ANR (resp., a G/N -AR). In particular, X/G is an ANR (resp., an AR).*

Recall that for a metric space (X, d) , the Hausdorff metric d_H on $\exp X$ is defined by the formula

$$d_H(C, D) = \max \left\{ \sup_{x \in D} \text{dist}(x, C), \sup_{y \in C} \text{dist}(y, D) \right\} \quad \text{for } C, D \in \exp X.$$

The topology generated by d_H is an invariant of the topology of X (it does not depend on d).

If G is a compact group and X is a metrizable G -space, then the formula

$$(g, A) \mapsto gA; \quad gA = \{ga \mid a \in A\}, \text{ for all } g \in G, A \in \exp X$$

defines a continuous G -action on $\exp X$; so $\exp X$ naturally becomes a G -space. In this case the complement $\exp_0 X = (\exp X) \setminus \{X\}$ is an open invariant subset of $\exp X$. Clearly, if d is a G -invariant metric on X , then d_H is a G -invariant metric on $\exp X$.

For the boundary of a set $A \subset X$ we will use the notation ∂A .

Throughout the paper we will use the following standard notation:

$$B^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + \dots + x_n^2 \leq 1\}, \text{ the Euclidean unit ball;}$$

$$S^{n-1} = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + \dots + x_n^2 = 1\}, \text{ the Euclidean unit sphere;}$$

$$\Delta^n = \{(t_0, \dots, t_n) \in \mathbb{R}^{n+1} \mid t_i \geq 0, t_0 + \dots + t_n = 1\}, \text{ the standard closed simplex;}$$

$$Q = \prod_{k=1}^{\infty} \{I_k \mid I_k = [0, 1]\}, \text{ the Hilbert cube.}$$

3. PROOF OF THEOREM 1.1 AND ITS COROLLARIES

We first prove the following equivariant version of Curtis' generalization [13, Theorem 1.6] of the well-known Wojdyslawski Theorem [35], which is just the implication (1) \implies (2) in Theorem 1.1.

Recall that a metric space X is *continuum-connected* if each pair of points in X is contained in a subcontinuum. X is *locally continuum-connected* if it has an open base of continuum-connected subsets [13].

Proposition 3.1. *Let G be a compact group and X a locally continuum-connected (resp., connected and locally continuum-connected) metrizable G -space. Then $\exp X$ is a G -ANR (resp., G -AR).*

Proof. We shall consider only the “ G -ANR” case. The “ G -AR” case is similar.

If $\exp X$ is a G -ANR, then it is also an ANR, and then by Curtis' theorem [13], X is locally continuum-connected.

Now assume that (Y, ρ) is a metric G -space with ρ an invariant metric on Y , A a closed invariant subset of Y , and $\varphi : A \rightarrow \exp X$ a G -map. By Curtis' theorem [13], $\exp X$ is an ANR. So φ has a continuous extension $f : U \rightarrow \exp X$ defined on a neighborhood U of A in Y . By compactness of G , there is an invariant neighborhood V of A contained in U . Set

$$F(y) = \bigcup_{g \in G} g^{-1} \varphi(gy) \quad \text{for every } y \in V.$$

We claim that the map $F : V \rightarrow \exp X$ is a well-defined continuous G -equivariant extension of φ . Indeed, for every $y \in V$ the set $\{g^{-1} \varphi(gy) \mid g \in G\}$ is a compact subset of $\exp X$ because it is the image of the continuous map $\alpha : G \rightarrow \exp X$, $\alpha(g) = g^{-1} \varphi(gy)$. Therefore, the union $\bigcup_{g \in G} g^{-1} \varphi(gy)$ is a compact subset of X , i.e., $F(y) \in \exp X$.

Further, if $t, g \in G$ and $h = gt$, then

$$F(ty) = \bigcup_{g \in G} g^{-1} \varphi(gty) = \bigcup_{h \in G} t(h^{-1} \varphi(hy)) = t \left(\bigcup_{h \in G} h^{-1} \varphi(hy) \right) = tF(y),$$

showing the equivariance of F . If $a \in A$, then $\varphi(ga) = f(ga)$ for all $g \in G$, and by the equivariance of f we will then have

$$F(a) = \bigcup_{g \in G} g^{-1} \varphi(ga) = \bigcup_{g \in G} g^{-1} f(ga) = \bigcup_{g \in G} g^{-1} g f(a) = \bigcup_{g \in G} f(a) = f(a),$$

showing that F extends f .

Let d be an invariant metric on X .

To see the continuity of F , we fix $y_0 \in V$ and $\varepsilon > 0$ arbitrary. By continuity of φ , for each $g \in G$ there is a $\delta_g > 0$ such that

$$d_H(\varphi(gy_0), \varphi(z)) < \varepsilon/2 \quad \text{whenever } z \in V \text{ and } \rho(z, gy_0) < 2\delta_g.$$

By compactness of the orbit $G(y_0)$, its open cover $\{O(gy_0, \delta_g) \mid g \in G\}$, where $O(z, r)$ stands for the open r -ball in X centered at z , admits a finite subcover

$$\{O(g_1 y_0, \delta_{g_1}), \dots, O(g_k y_0, \delta_{g_k})\}.$$

Let $\delta = \min\{\delta_{g_1}, \dots, \delta_{g_k}\}$. We are going to check that then $d_H(F(y), F(y_0)) < \varepsilon$ whenever $y \in V$ and $\rho(y, y_0) < \delta$.

Indeed, for every $g \in G$, gy_0 belongs to an element of the cover

$$\{O(g_1y_0, \delta_{g_1}), \dots, O(g_ky_0, \delta_{g_k})\}.$$

Without loss of generality, one can assume that $gy_0 \in O(g_1y_0, \delta_{g_1})$. Then for every $y \in O(y_0, \delta)$ we have

$$\rho(gy, g_1y_0) \leq \rho(gy, gy_0) + \rho(gy_0, g_1y_0) = \rho(y, y_0) + \rho(gy_0, g_1y_0) \leq \delta + \delta_{g_1} \leq 2\delta_{g_1}.$$

Thus, $gy_0, gy \in O(g_1y_0, 2\delta_{g_1})$, which implies that $d_H(\varphi(gy), \varphi(gy_0)) < \varepsilon$. By the invariance of d_H , this yields

$$d_H(g^{-1}\varphi(gy), g^{-1}\varphi(gy_0)) = d_H(\varphi(gy), \varphi(gy_0)) < \varepsilon \text{ for all } g \in G,$$

which in turn implies that

$$d_H\left(\bigcup_{g \in G} g^{-1}\varphi(gy), \bigcup_{g \in G} g^{-1}\varphi(gy_0)\right) < \varepsilon.$$

Thus, $d_H(F(y), F(y_0)) < \varepsilon$ for all $y \in O(y_0, \delta)$, proving the continuity of F at the point y_0 . \square

Remark 3.2. Recall that a hyperspace $\mathcal{E} \subset \exp X$ that satisfies the following condition is called an *inclusion hyperspace*: if $B \in \exp X$ and $A \in \mathcal{E}$ is such that $A \subset B$, then $B \in \mathcal{E}$. As is clear from the proof, Proposition 3.1 remains true also for any invariant inclusion hyperspace \mathcal{E} instead of $\exp X$.

(2) \implies (3) follows directly from Theorem 2.7.

Since every ANR (resp., AR) is locally continuum-connected (resp., connected and locally continuum-connected), the implication (3) \implies (1) follows from the following result.

Proposition 3.3. *Let G be a compact group and X a metric G -space. Then:*

- (1) *X is connected iff $(\exp X)/G$ is connected; and*
- (2) *if in addition G is a Lie group, then X is locally continuum-connected iff $(\exp X)/G$ is locally path-connected.*

For the proof we need the following three lemmas.

Lemma 3.4. *Let G be a compact group, and X a G -space containing a connected invariant subset (e.g., G connected or $X[G] \neq \emptyset$). Then X is connected iff X/G is connected.*

Proof. Only the “if” part requires a proof. Let $C \subset X$ be an invariant connected set. Assume the contrary, that X/G is connected and X is not connected. Then $X = A_1 \cup A_2$ where the A_i , $i = 1, 2$, are nonempty, disjoint, closed-open subsets of X . Since C is connected, one (and only one) of the sets A_i , $i = 1, 2$, contains C . Suppose $C \subset A_1$. By compactness of G , the orbit map $p : X \rightarrow X/G$ is open and closed. Hence the set $p(A_2)$ is a nonempty open-closed subset of X/G . Besides, by the invariance of C the set $p(A_2)$ is disjoint from $p(C)$; so $p(A_2) \neq X/G$, which contradicts the connectedness of X/G . \square

Lemma 3.5. *Let G be a compact Lie group, and X a G -space containing a path-connected invariant subset (e.g., G connected or $X[G] \neq \emptyset$). Then X is path-connected iff X/G is path-connected.*

Proof. Only the "if" part requires a proof. Let $C \subset X$ be an invariant path-connected set. It suffices to show that each point $x \in X$ can be joined with a point of C . Let $c \in C$ and let $l : [0, 1] \rightarrow X/G$ be a path with $l(0) = p(x)$, $l(1) = p(c)$, where $p : X \rightarrow X/G$ is the orbit projection. By [11, Ch. II, Theorem 6.2] there is a lifting $l' : [0, 1] \rightarrow X$, $pl' = l$. Since $l'(0)$ belongs to the orbit of x , there is $g \in G$ such that $x = gl'(0)$. Then the path $gl' : [0, 1] \rightarrow X$, $(gl')(t) = gl'(t)$, connects x and $gl'(1)$. But $gl'(1) \in C$, because $l'(1) \in C$ and C is invariant. This completes the proof. \square

Lemma 3.6. *Let G be a compact Lie group. Then a G -space X is locally path-connected iff X/G is locally path-connected.*

Proof. Only the "if" part requires a proof. Let $x \in X$ and let U be a neighborhood of x . Since the action of G is continuous and G is locally path-connected, one can choose a path-connected neighborhood O of the unity in G and a neighborhood V of x such that $OV = \{gv \mid g \in O, v \in V\} \subset U$. Without loss of generality, one can assume that V is G_x -invariant. Since G is a compact Lie group, there is a G_x -slice R containing the point x [26, Corollary 1.7.17]. Then the set $T = R \cap V$ is also a G_x -slice containing x . Since X/G is locally path-connected, the neighborhood $p(G(T))$ of the point $p(x)$ contains a path-connected neighborhood \tilde{Y} of $p(x)$ in X/G . Let $Y = p^{-1}(\tilde{Y})$. Then Y is a G -invariant neighborhood of the orbit $G(x)$ in X . Clearly, the set $S = T \cap Y$ contains x and is a global G_x -slice for the G -space Y . Therefore the orbit spaces S/G_x and $Y/G = \tilde{Y}$ are homeomorphic [26, Proposition 1.7.6]; in particular, S/G_x is path-connected. By Lemma 3.5, now S is path-connected as well. Consequently, OS is path-connected. Since OS is an open neighborhood of x and $OS \subset OT \subset OV \subset U$, we are done. \square

Proof of Proposition 3.3. 1. It is well known that X is connected iff $\exp X$ is so; this is proved, for instance, in [23, Proposition 5.3.10] for X compact, but the same proof is valid for noncompact X as well. Hence, it remains to show that the connectedness of $(\exp X)/G$ implies that of $\exp X$. If $(\exp X)/G$ is connected, then the invariant subset Γ of $\exp X$ consisting of all invariant, compact subsets of X is connected. Indeed, the continuous surjection $\exp X \rightarrow \exp(X/G)$ induced by the orbit map $X \rightarrow X/G$ is invariant, and hence, it induces a continuous surjection $(\exp X)/G \rightarrow \exp(X/G)$. Therefore, if $(\exp X)/G$ is connected, then $\exp(X/G)$ is so, and since Γ is homeomorphic to $\exp(X/G)$, it is connected too. Thus, Γ is an invariant, connected subset of the G -space $\exp X$, and the connectedness of $(\exp X)/G$ implies, by Lemma 3.4, the connectedness of $\exp X$.

2. Respectively, X is locally continuum-connected iff $\exp X$ is locally path-connected (see [13, Lemma 1.4] and the final part of the proof of [13, Theorem 1.6]). By Lemma 3.6, $\exp X$ is locally path-connected iff $(\exp X)/G$ is locally path-connected, and the proof is complete. Thus, Theorem 1.1 is completely proved.

Corollary 3.7. *Let G be a compact, metrizable, locally path-connected group, and $H \subset G$ a closed subgroup. Then $(\exp(G/H))/G$ is an ANR. If in addition G/H is connected, then $(\exp(G/H))/G$ is an AR.*

Proof. Under the hypothesis the coset space G/H is locally path-connected, and hence, locally continuum-connected. Besides, since the quotient map $G \rightarrow G/H$ is perfect, G/H is metrizable too (see, e.g., [15, Section XI, Theorem 5.2(3)]). Then

by Proposition 3.3, $\exp(G/H) \in G\text{-ANR}$; if in addition G/H is connected, then $\exp(G/H) \in G\text{-AR}$. Now the result follows immediately from Theorem 2.7. \square

Since every compact Lie group G is locally path-connected (moreover, it is a $G\text{-ANR}$ [26, Corollary 1.6.7]), we get from Corollary 3.7 the following positive answer to the first question of West's Problem:

Corollary 3.8. *Let G be a compact Lie group. Then $(\exp G)/G$ is an ANR. If in addition G is connected, then $(\exp G)/G$ is an AR.*

In combination with [17, Corollary 2], Theorem 1.1 yields also the following fact.

Corollary 3.9. *Let G be a finite group acting on a nondegenerate Peano continuum X . Then the orbit space $(\exp X)/G$ is a Hilbert cube.*

4. THE G -NERVE

In this section we develop a necessary technique involving the notion of a G -nerve. The results proved here will be applied in the next section.

Following Matumoto [22], we define the G -nerve $\mathcal{N}(\mathcal{U})$ of a G -normal cover \mathcal{U} .

Let G be a compact Lie group and $\mathcal{H} = \{H_\mu \mid \mu \in M\}$ a family of closed subgroups of G . The Milnor join \mathcal{J} of the family of cosets $\{G/H_\mu \mid \mu \in M\}$ is defined as follows. Let \mathcal{J}' be the following set:

$$\left\{ (t_\mu, g_\mu H_\mu)_{\mu \in M} \in \prod_{\mu \in M} (I \times G/H_\mu) \mid t_\mu \neq 0 \text{ for only finite } \mu\text{'s and } \sum_{\mu \in M} t_\mu = 1 \right\}.$$

We let

$$(t_\mu, g_\mu H_\mu)_{\mu \in M} \sim (s_\mu, h_\mu H_\mu)_{\mu \in M}$$

iff $t_\mu = s_\mu$ for all $\mu \in M$, and $g_\mu H_\mu = h_\mu H_\mu$ whenever $t_\mu \neq 0$. Then \sim is an equivalence relation on \mathcal{J}' , and we shall denote by \mathcal{J} the quotient set \mathcal{J}' / \sim .

In what follows we shall use the convention $\sum_{\mu \in M} t_\mu g_\mu H_\mu$ for the equivalence class of the point $(t_\mu, g_\mu H_\mu)_{\mu \in M}$. The numbers t_μ are called barycentric coordinates of the point $\sum_{\mu \in M} t_\mu g_\mu H_\mu \in \mathcal{J}$.

For any finite subset $\{\mu_0, \dots, \mu_n\} \subset M$, we consider the following subset of \mathcal{J} :

$$G/H_{\mu_0} * \dots * G/H_{\mu_n} = \left\{ \sum_{\mu \in M} t_\mu g_\mu H_\mu \in \mathcal{J} \mid t_\mu = 0 \text{ for all } \mu \notin \{\mu_0, \dots, \mu_n\} \right\}.$$

Observe that each $G/H_{\mu_0} * \dots * G/H_{\mu_n}$ with its quotient topology is a compact metrizable space, which is called the Milnor join of the spaces $G/H_{\mu_0}, \dots, G/H_{\mu_n}$ (see [24]).

We topologize \mathcal{J} by the weak topology with respect to the family of all finite subjoins, i.e., a set $U \subset \mathcal{J}$ is open in \mathcal{J} whenever $U \cap (G/H_{\mu_0} * \dots * G/H_{\mu_n})$ is open in $G/H_{\mu_0} * \dots * G/H_{\mu_n}$ for any finite subjoin $G/H_{\mu_0} * \dots * G/H_{\mu_n} \subset \mathcal{J}$. It is easy to check that \mathcal{J} becomes a G -space if we define the action of G as follows:

$$g \left(\sum_{\mu \in M} t_\mu g_\mu H_\mu \right) = \sum_{\mu \in M} t_\mu g g_\mu H_\mu, \quad g \in G.$$

Next, if $g_{\mu_0}H_{\mu_0} \in G/H_{\mu_0}, \dots, g_{\mu_n}H_{\mu_n} \in G/H_{\mu_n}$ are fixed elements, then we will denote by $\langle g_{\mu_0}H_{\mu_0}, \dots, g_{\mu_n}H_{\mu_n} \rangle$ the subspace

$$\left\{ \sum_{\mu \in M} t_{\mu} g'_{\mu} H_{\mu} \in \mathcal{J} \mid t_{\mu} = 0 \text{ for } \mu \notin \{\mu_0, \dots, \mu_n\} \right. \\ \left. \text{and } g'_{\mu_i} H_{\mu_i} = g_{\mu_i} H_{\mu_i} \text{ for } 0 \leq i \leq n \right\}$$

of $G/H_{\mu_0} * \dots * G/H_{\mu_n}$, which is called an n -cell.

Let G be a compact Lie group and X a G -space. For each index $\mu \in M$, let H_{μ} be a closed subgroup of G , and let S_{μ} be an H_{μ} -slice in X . Then the family

$$\mathcal{U} = \{(gS_{\mu}, H_{\mu}) \mid g \in G, \mu \in M\}$$

consisting of tubular slice-sets gS_{μ} with companion groups H_{μ} is called a G -normal cover of X if the family of open tubes $\{G(S_{\mu}) \mid \mu \in M\}$ covers X and there exists a locally finite invariant partition of unity $\{\varphi_{\mu} : X \rightarrow [0, 1] \mid \mu \in M\}$ subordinated to \mathcal{U} , i.e., each φ_{μ} is an invariant function with $\varphi_{\mu}^{-1}((0, 1]) \subset G(S_{\mu})$ and the supports $\{\varphi_{\mu}^{-1}((0, 1]) \mid \mu \in M\}$ constitute a locally finite family.

Let $\tilde{\mathcal{N}}(\mathcal{U})$ be the ordinary nerve of the invariant cover $\{G(S_{\mu}) \mid \mu \in M\}$. In the sequel we will denote by $\langle \mu_0, \dots, \mu_n \rangle$ the simplex of $\tilde{\mathcal{N}}(\mathcal{U})$ constituted by the sets $G(S_{\mu_0}), \dots, G(S_{\mu_n})$. Let $f_{\mu} : G(S_{\mu}) \rightarrow G/H_{\mu}$ be the corresponding G -map with $f_{\mu}^{-1}(eH_{\mu}) = S_{\mu}$ (see Slice Theorem 2.2). For any simplex $L = \langle \mu_0, \dots, \mu_n \rangle \in \tilde{\mathcal{N}}(\mathcal{U})$, we define the following subset of the product $\prod_{i=0}^n G/H_{\mu_i}$:

$$F_L = \left\{ (f_{\mu_0}(x), \dots, f_{\mu_n}(x)) \mid x \in \bigcap_{i=0}^n G(S_{\mu_i}) \right\}.$$

It follows from the equivariance of f_{μ_i} that F_L is an invariant subset of the G -space $\prod_{i=0}^n G/H_{\mu_i}$.

Denote by \mathcal{F} the family of all these sets F_L . Let $\Delta(L, F_L)$ be the subset of the finite subjoin $G/H_{\mu_0} * \dots * G/H_{\mu_n}$ of \mathcal{J} consisting of all those points $\sum_{i=0}^n t_i g_i H_{\mu_i}$ for which $(g_0 H_{\mu_0}, \dots, g_n H_{\mu_n}) \in F_L$ and $(t_0, \dots, t_n) \in \Delta^n$. Clearly, $\Delta(L, F_L)$ is invariant in $G/H_{\mu_0} * \dots * G/H_{\mu_n}$, and hence, in \mathcal{J} .

We call $\Delta(L, F_L)$ a G - n -simplex over the simplex L along the set F_L . The homogeneous G -spaces $G/H_{\mu_0}, \dots, G/H_{\mu_n}$ are called G -vertices of $\Delta(L, F_L)$.

The G -nerve of the cover \mathcal{U} is, by definition, the union

$$\mathcal{N}(\mathcal{U}) = \bigcup \left\{ \Delta(L, F_L) \mid L \in \tilde{\mathcal{N}}(\mathcal{U}), F_L \in \mathcal{F} \right\}$$

equipped with the topology induced from \mathcal{J} . It is not difficult to check that the topology of $\mathcal{N}(\mathcal{U})$ is the weak one with respect to its closed, invariant cover $\{\Delta(L, F_L) \mid L \in \tilde{\mathcal{N}}(\mathcal{U}), F_L \in \mathcal{F}\}$. That is to say, a subset $W \subset \mathcal{N}(\mathcal{U})$ is open iff $W \cap \Delta(L, F_L)$ is open in $\Delta(L, F_L)$ for each G -simplex $\Delta(L, F_L)$ in $\mathcal{N}(\mathcal{U})$. Since $\mathcal{N}(\mathcal{U})$ is an invariant subset of \mathcal{J} , it becomes a G -space with respect to the action induced from \mathcal{J} .

The G - n -skeleton $\mathcal{N}(\mathcal{U})^n$ of $\mathcal{N}(\mathcal{U})$ is defined to be the union of all G - k -simplexes in $\mathcal{N}(\mathcal{U})$ with $k \leq n$.

If gH_λ is a vertex of the G -nerve $\mathcal{N}(\mathcal{U})$, then its star $St(gH_\lambda, \mathcal{N}(\mathcal{U}))$ is defined to be the union of all cells for which gH_λ is a vertex. The G -carrier of a point $x \in \mathcal{N}(\mathcal{U})$ is defined to be the smallest G -simplex of $\mathcal{N}(\mathcal{U})$ containing x . The cell in the G -carrier that contains x is called the *carrier* of x .

Recall that a cover \mathcal{U} of a space X is called a star-refinement of a cover \mathcal{V} whenever for every $U \in \mathcal{U}$ there exists an element $V \in \mathcal{V}$ that contains the star $St(U, \mathcal{U})$ of U with respect of \mathcal{U} ; here $St(U, \mathcal{U}) = \{W \in \mathcal{U} \mid W \cap U \neq \emptyset\}$.

Lemma 4.1. *Let X be a paracompact G -space. Then for each open cover \mathcal{V} of X there exists a G -normal cover $\mathcal{U} = \{(gS_\lambda, H_\lambda) \mid g \in G, \lambda \in \Lambda\}$ of X such that \mathcal{U} is a star-refinement of \mathcal{V} .*

Proof. Since X is paracompact, one can choose open covers \mathcal{U}_1 and \mathcal{U}_2 of X such that \mathcal{U}_1 is a star-refinement of \mathcal{U}_2 and \mathcal{U}_2 is a star-refinement of \mathcal{V} .

Let us denote by U the subset of $X \times X$ consisting of all those pairs (x, y) such that there exists an element $O \in \mathcal{U}_1$ that contains both x and y . Clearly U is an open neighborhood of the diagonal $\Delta \subset X \times X$. By compactness of G , there is an invariant neighborhood V of Δ in $X \times X$ such that $V \subset U$. Define

$$\mathcal{W} = \{V[x] \mid x \in X\}, \quad \text{where} \quad V[x] = \{z \in X \mid (x, z) \in V\}.$$

Then \mathcal{W} is an open G -cover of X , and $V[x] \subset U[x] = St(x, \mathcal{U}_1)$ for each $x \in X$. Since $St(x, \mathcal{U}_1)$ is contained in an element of \mathcal{U}_2 , we infer that \mathcal{W} is a refinement of \mathcal{U}_2 , and hence, a star-refinement of \mathcal{U} .

Next, we fix on each orbit $G(x) \subset X$ a point, say $x \in G(x)$, and choose an element $W_x \in \mathcal{W}$ such that $x \in W_x$. By the continuity of the action of G on X there exist a neighborhood O_x of the unity in G and a G_x -invariant neighborhood N_x of x in X such that $O_x N_x \subset W_x$. By Slice Theorem 2.2, there exists a G_x -slice Q_x such that $x \in Q_x$. Then the set $S_x = Q_x \cap N_x$ is also a G_x -slice, and $x \in S_x \subset N_x$. Hence $O_x S_x \subset W_x$. We define \mathcal{U} to be the totality of all these slice-sets (gS_x, G_x) , $g \in G$, $G(x) \in X/G$. Since the orbit map $X \rightarrow X/G$ is closed, we see that X/G is paracompact too [15, Section VIII, Theorem 2.4]. This implies that the invariant cover $\{G(S_x)\}$ admits a locally finite partition of unity subordinated to \mathcal{U} , and hence, \mathcal{U} is an open G -normal cover. Since $gS_x \subset gW_x$ and $gW_x \in \mathcal{W}$, we conclude that \mathcal{U} is a refinement of \mathcal{W} , and since \mathcal{W} is a star-refinement of \mathcal{V} , we infer that \mathcal{U} is a star-refinement of \mathcal{V} . \square

Lemma 4.2. *Let Y be a G -space, and let $\mathcal{U} = \{(gS_\mu, H_\mu) \mid g \in G, \mu \in M\}$ be a G -normal cover of Y . Then for each locally finite invariant partition of unity subordinated to \mathcal{U} , there exists a G -map $p: Y \rightarrow \mathcal{N}(\mathcal{U})$ such that $p^{-1}(St(gH_\mu, \mathcal{N}(\mathcal{U}))) \subset G(S_\mu)$ for any $g \in G$ and $\mu \in M$, where*

$$St(gH_\mu, \mathcal{N}(\mathcal{U})) = \left\{ \sum_{\lambda \in M} t_\lambda g_\lambda H_\lambda \in \mathcal{N}(\mathcal{U}) \mid t_\mu > 0, g_\mu = g \right\}.$$

Proof. Let $\{\varphi_\mu: Y \rightarrow [0, 1] \mid \mu \in M\}$ be a locally finite invariant partition of unity subordinated to \mathcal{U} , i.e., $\varphi_\mu^{-1}((0, 1]) \subset G(S_\mu)$ for all $\mu \in M$. Then we define the canonical G -map $p: Y \rightarrow \mathcal{N}(\mathcal{U})$ as follows.

Since $\{\varphi_\mu: Y \rightarrow [0, 1] \mid \mu \in M\}$ is locally finite, for each $y \in Y$ there are only a finite number of indices, say μ_0, \dots, μ_n , such that $\varphi_{\mu_i}(y) \neq 0$, $i = 0, \dots, n$. Let $\langle f_{\mu_0}(y), \dots, f_{\mu_n}(y) \rangle$ be the corresponding n -cell in $G/H_{\mu_0} * \dots * G/H_{\mu_n}$. Then by

definition, $p(y)$ is the point of $\langle f_{\mu_0}(y), \dots, f_{\mu_n}(y) \rangle$ with the barycentric coordinates $\varphi_{\mu_0}(y), \dots, \varphi_{\mu_n}(y)$, i.e.,

$$p(y) = \sum_{i=0}^n \varphi_{\mu_i}(y) f_{\mu_i}(y).$$

We claim that p is continuous. For, let $y_0 \in Y$ be an arbitrary point. Using the local finiteness of the partition of unity $\{\varphi_{\mu} \mid \mu \in M\}$, we take a neighborhood V of y_0 in Y such that only for a finite number of indices μ_0, \dots, μ_m is $\varphi_{\mu_i}(y) \neq 0$ for $y \in V$. Then

$$p(y) = \sum_{i=0}^m \varphi_{\mu_i}(y) f_{\mu_i}(y) \quad \text{for all } y \in V.$$

Now the continuity of p in V follows from the continuity of the maps f_{μ_i} and φ_{μ_i} , $i = 0, \dots, m$, in V . \square

For a space X we will denote by $\mathcal{F}_n(X)$ the subset of $\exp X$ that consists of all those sets $A \subset X$ that have at most n elements. By $\mathcal{F}_{\infty}(X)$ we shall denote the union $\bigcup_{n=1}^{\infty} \mathcal{F}_n(X)$.

Lemma 4.3. *Let P be a polyhedron and P^1 its 1-dimensional skeleton. Then there is a continuous map $\xi : P \rightarrow \mathcal{F}_{\infty}(P^1)$ such that*

- (1) $\xi(z) = \{z\}$ for all $z \in P^1$, and
- (2) if τ is the carrier of $x \in P$ and $\dim \tau = n$, then $\xi(x)$ is contained in the 1-skeleton of τ and contains at most 3^{n-1} points.

Proof. Let P^n , $n \geq 1$, denote the n -skeleton of P , and let $\xi_1 : P^1 \rightarrow P^1$ be the identity map. From the proof of [23, Proposition 8.4.2], we get the following fact.

Claim. For every $n \geq 1$, there is a continuous map $\xi_n : P^n \rightarrow \mathcal{F}_{\infty}(P^1)$ such that

- a. if τ is the carrier of $x \in P$ and $\dim \tau = n$, then $\xi_n(x)$ is contained in the 1-skeleton of τ and contains at most 3^{n-1} points, and
- b. ξ_{n+1} extends ξ_n for all $n \geq 1$.

Now we define the required map ξ to be equal to ξ_n on the n -dimensional skeleton P^n . \square

In the next lemma, for a given simplex $L = \langle \mu_0, \dots, \mu_n \rangle \subset \tilde{\mathcal{N}}(\mathcal{U})$, a given n -cell $\langle g_{\mu_0} H_{\mu_0}, \dots, g_{\mu_n} H_{\mu_n} \rangle \subset \mathcal{N}(\mathcal{U})$ and the corresponding G - n -simplex $\sigma = \Delta(L, F_L)$, we shall use the following notation:

$$\partial \langle g_{\mu_0} H_{\mu_0}, \dots, g_{\mu_n} H_{\mu_n} \rangle = \left\{ \sum_{i=0}^n t_i g_{\mu_i} H_{\mu_i} \mid (t_0, \dots, t_n) \in \partial \Delta^n \right\},$$

where we use the same notation $\partial \Delta^n$ for the ordinary boundary of the standard simplex Δ^n . Correspondingly,

$$\partial \sigma = \left\{ \sum_{i=0}^n t_i g_{\mu_i} H_{\mu_i} \mid (t_0, \dots, t_n) \in \partial \Delta^n, (g_{\mu_0} H_{\mu_0}, \dots, g_{\mu_n} H_{\mu_n}) \in F_L \right\}.$$

In what follows we shall need the following equivariant version of Lemma 4.3:

Lemma 4.4. *Let $\mathcal{N}(\mathcal{U})$ be a G -nerve and Γ its G -1-dimensional skeleton. Then there is a G -map $R : \mathcal{N}(\mathcal{U}) \rightarrow \mathcal{F}_{\infty}(\Gamma)$ such that*

- (1) $R(z) = \{z\}$ for all $z \in \Gamma$, and

- (2) if s is the carrier of $x \in \mathcal{N}(\mathcal{U})$, then $R(x)$ is contained in the 1-skeleton of s . More precisely, if $\dim s = n$, then $R(x) \in \mathcal{F}_{3n-1}(s)$.

Proof. We are going to apply Lemma 4.3 above. In our case P is the polyhedron accompanying the G -nerve $\mathcal{N}(\mathcal{U})$, i.e., the polyhedron $\tilde{\mathcal{N}}(\mathcal{U})$. Let K be its 1-skeleton and $\xi : P \rightarrow \mathcal{F}_\infty(K)$ the continuous map from Lemma 4.3.

Let $x \in \mathcal{N}(\mathcal{U})$ and let the G -simplex

$$\sigma = \Delta(L, F_L) \subset G/H_{\mu_0} * \cdots * G/H_{\mu_n}$$

be the G -carrier of x . Then $x = \sum_{i=0}^n t_i g_{\mu_i} H_{\mu_i}$, where $s = \langle g_{\mu_0} H_{\mu_0}, \dots, g_{\mu_n} H_{\mu_n} \rangle$ is the carrier of x .

Define the map $R : \mathcal{N}(\mathcal{U}) \rightarrow \mathcal{F}_\infty(\Gamma)$ by setting

$$\begin{aligned} R(x) &= R\left(\sum_{i=0}^n t_i g_{\mu_i} H_{\mu_i}\right) = \xi(t_0, \dots, t_n) \cdot (g_{\mu_0} H_{\mu_0}, \dots, g_{\mu_n} H_{\mu_n}) \\ &= \left\{ \sum_{i=0}^n u_i g_{\mu_i} H_{\mu_i} \mid (u_0, \dots, u_n) \in \xi(t_0, \dots, t_n) \right\}. \end{aligned}$$

Since $\xi(t_0, \dots, t_n)$ belongs to $\mathcal{F}_{3n-1}(P^1)$, we see that $R(x)$ belongs to $\mathcal{F}_{3n-1}(s) \subset \mathcal{F}_{3n-1}(\Gamma)$. Continuity and equivariance of R are evident from the definition of R . Properties (1) and (2) follow from the analogous properties in Lemma 4.3. \square

5. PROOFS OF THEOREMS 1.3 AND 1.4

We shall give a sequence of lemmas culminating in proofs of Theorems 1.3 and 1.4.

In this section d will always denote the Euclidean metric on \mathbb{R}^n .

By $\mathcal{P}(n)$ we will denote the subset of $L(n)$ consisting of all compact convex bodies A such that the contact set $\partial A \cap \partial B^n$ has an empty interior in the boundary sphere $S^{n-1} = \partial B^n$.

Lemma 5.1. *For each $\varepsilon > 0$ and each body $E \in L_0(n)$ there exists a body $D \in \mathcal{P}(n)$ such that $d_H(E, D) < \varepsilon$ and the $O(n)$ -stabilizer $O(n)_E$ of E coincides with the $O(n)$ -stabilizer $O(n)_D$ of D .*

Proof. Let $r : \mathcal{B}(n) \rightarrow L(n)$ be the $O(n)$ -equivariant retraction from Theorem 2.3.

Because of compactness of $L(n)$ (Theorem 2.4), one can find a real $0 < \delta < \varepsilon/2$ such that $d_H(r(A), A) < \varepsilon/2$ for all A belonging to the δ -neighborhood of $L(n)$ in $\mathcal{B}(n)$, where d_H denotes the Hausdorff metric on $\mathcal{B}(n)$.

Let K be the stabilizer $O(n)_E$. It follows from Slice Theorem 2.2 that there is a real $0 < \eta < \delta$ such that the inequality $d_H(E, X) < \eta$ for $X \in L_0(n)$ implies that the stabilizer $O(n)_X$ is conjugate to a subgroup of K , i.e., $(O(n)_X) \preceq (K)$.

Choose a centrally symmetric, convex polyhedron $P \subset \mathbb{R}^n$ with a nonempty interior, such that $d_H(E, P) < \eta$, $P \subset E$ and all the vertices p_1, \dots, p_k of P lie on the boundary ∂E . Then the convex hull

$$M = \operatorname{conv}(K(p_1) \cup \cdots \cup K(p_k))$$

is a centrally symmetric, compact, convex, K -invariant subset of \mathbb{R}^n . Since it contains P , we see that M has a nonempty interior in \mathbb{R}^n , and so $M \in \mathcal{B}(n)$.

We claim that the boundary ∂M does not contain an $(n-1)$ -dimensional elliptic domain, i.e., an open subset $V \subset \partial M$ which is at the same time an open subset

of some $(n - 1)$ -dimensional ellipsoid surface lying in \mathbb{R}^n . It suffices to show that none of the orbits $K(p_i)$, $i = 1, \dots, k$, contains an $(n - 1)$ -dimensional elliptic domain. Assume the contrary, that some $K(p_i)$ contains an $(n - 1)$ -dimensional elliptic domain V . Since $K(p_i)$ lies on the $(n - 1)$ -dimensional sphere $\partial B(0, \|p_i\|)$ centered at the origin and having the radius $\|p_i\|$, then V should be, in fact, a domain of the sphere $\partial B(0, \|p_i\|)$. Since $K(p_i)$ is homogeneous and compact, we conclude that there are finitely many open subsets V_1, \dots, V_n of $K(p_i)$ such that $K(p_i) = V_1 \cup \dots \cup V_n$, where each V_j is homeomorphic to V . Next, by the Domain Invariance Theorem (see, e.g., [27, Ch. 4, Section 7, Theorem 16]), each V_j should be open in the sphere $\partial B(0, \|p_i\|)$, and hence, the union $V_1 \cup \dots \cup V_n$ is open in $\partial B(0, \|p_i\|)$. But $V_1 \cup \dots \cup V_n$ is also closed in $\partial B(0, \|p_i\|)$, because it is equal to $K(p_i)$. Now, by connectedness of $\partial B(0, \|p_i\|)$, it then follows that $K(p_i)$ must be the whole sphere $\partial B(0, \|p_i\|)$. Consequently, K acts transitively on the sphere $\partial B(0, \|p_i\|)$, and hence on the unit sphere S^{n-1} . This contradiction proves the claim.

In particular, the contact set of M , which is by definition the intersection of ∂M with the boundary of the Löwner ellipsoid $l(M)$, also does not contain an elliptic domain.

Now consider the body $D = r(M) \in L(n)$. Since $D = T(M)$ for some linear nondegenerate operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the contact set $\partial D \cap \partial B^n$ is just the image under T of the contact set $\partial M \cap \partial(l(M))$. Consequently, $\partial D \cap \partial B^n$ has an empty interior in the sphere $S^{n-1} = \partial B^n$, and hence, $D \in \mathcal{P}(n)$. Since $d_H(P, E) < \eta$ and $P \subset M \subset E$, we see that $d_H(M, E) < \eta < \delta$. Consequently, $d_H(D, E) < \varepsilon$, which implies in turn that $(O(n)_D) \preceq (K)$. Since M is K -invariant, we have $K \subset O(n)_M$, and since r is $O(n)$ -equivariant, $O(n)_M \subset O(n)_{r(M)} = O(n)_D$. Thus $K \subset O(n)_D$, which implies in combination with $(O(n)_D) \preceq (K)$ that $O(n)_D = K$. \square

The following lemma is the key in the proof of Theorems 1.3 and 1.4:

Lemma 5.2 (Key lemma). *Let $\varepsilon > 0$, and let \mathcal{V}_ε be the ε -cover of $L_0(n)$. Then there exist a G -normal cover $\mathcal{U} = \{(gS_\lambda, H_\lambda) \mid g \in G, \lambda \in \Lambda\}$ of $L_0(n)$ and G -maps $p : L_0(n) \rightarrow \mathcal{N}(\mathcal{U})$, $\Phi : \mathcal{N}(\mathcal{U}) \rightarrow \mathcal{P}(n)$ such that*

- (1) *for each $gS_\lambda \in \mathcal{U}$ there is an element $V(g, \lambda) \in \mathcal{V}_{\varepsilon/2}$ with $gS_\lambda \subset V(g, \lambda)$ and $\Phi(\text{St}(gH_\lambda, \mathcal{N}(\mathcal{U}))) \subset V(g, \lambda)$, and*
- (2) *the composition Φp is ε -close to the identity map of $L_0(n)$.*

Proof. By Lemma 4.1, there is a G -normal cover

$$\mathcal{U} = \{(gS_\lambda, H_\lambda) \mid g \in G, \lambda \in \Lambda\}$$

of $L_0(n)$ that is a star-refinement of $\mathcal{V}_{\varepsilon/4}$. Fix an invariant locally finite partition of unity $\{\varphi_\lambda\}_{\lambda \in \Lambda}$ subordinated to \mathcal{U} . Let $p : L_0(n) \rightarrow \mathcal{N}(\mathcal{U})$ be the G -map corresponding to $\{\varphi_\lambda\}_{\lambda \in \Lambda}$ (see Lemma 4.2).

For every $gS_\lambda \in \mathcal{U}$ we choose an element $V(g, \lambda) \in \mathcal{V}_{\varepsilon/2}$ such that

$$(5.1) \quad \text{St}(gS_\lambda, \mathcal{U}) \subset V(g, \lambda)/2,$$

where $V(g, \lambda)/2 \in \mathcal{V}_{\varepsilon/4}$ denotes the open ball in $L_0(n)$ concentric with $V(g, \lambda)$ and having half the radius of $V(g, \lambda)$.

Now we define the map $\Phi : \mathcal{N}(\mathcal{U}) \rightarrow \mathcal{P}(n)$ as follows.

First we define a G -map $q : \Gamma \rightarrow \mathcal{P}(n)$, where Γ is the G -1-skeleton of $\mathcal{N}(\mathcal{U})$.

For each G -vertex $G/H_\lambda \in \mathcal{N}(\mathcal{U})$ with H_λ the subgroup corresponding to the H_λ -slice $S_\lambda \in \mathcal{U}$, we select a point $X_\lambda \in S_\lambda$ such that the stabilizer G_{X_λ} coincides with the group H_λ .

By Lemma 5.1 above, we choose a body $A_\lambda \in \mathcal{P}(n)$, $\varepsilon/4$ -close to X_λ and having the stabilizer $G_{A_\lambda} = H_\lambda$.

Define the G -map $q : G/H_\lambda \rightarrow G(A_\lambda) \subset \mathcal{P}(n)$ by setting

$$q(gH_\lambda) = gA_\lambda \text{ for every } gH_\lambda \in G/H_\lambda.$$

The inclusion $H_\lambda \subset G_{A_\lambda}$ guarantees that q is a well-defined G -map. We aim to extend q equivariantly to the G -1-skeleton Γ of $\mathcal{N}(\mathcal{U})$. It suffices to extend q over each G -1-simplex of Γ . Let Δ be a G -1-simplex of Γ with the G -vertices G/H_λ and G/H_μ . We first define two G -maps $s_1 : \Delta \rightarrow \mathcal{B}(n)$ and $s_2 : \Delta \rightarrow \mathcal{B}(n)$. Indeed, let $y = tg_0H_\lambda + (1 - t)g_1H_\mu$ be an arbitrary point of Δ , where $g_0, g_1 \in G$ and $g_0S_\lambda \cap g_1S_\mu \neq \emptyset$.

We set

$$s_1(y) = tg_0A_\lambda + (1 - t)g_1A_\mu \quad \text{and} \quad s_2(y) = g_0A_\lambda.$$

Recall that here $tW + (1 - t)Z$ means the Minkowski convex combination of the convex sets $W, Z \subset \mathbb{R}^n$, i.e.,

$$tW + (1 - t)Z = \{tw + (1 - t)z \mid w \in W, z \in Z\}.$$

Clearly s_1 and s_2 are continuous G -maps from Δ to $\mathcal{B}(n)$.

Define $q'(y)$ to be the convex hull of the union $s_1(y) \cup s_2(y)$.

Since $s_1(y) \cup s_2(y)$ depends continuously upon $y \in \Delta$, the continuity of q' now follows from the continuity of the convex hull operator (see [31, Theorem 2.7.4(iv)]).

Analogously, let $q''(y)$ be the convex hull of the union

$$g_1A_\mu \cup (tg_0A_\lambda + (1 - t)g_1A_\mu).$$

For the same reason, $q''(y)$ depends continuously upon $y \in \Delta$.

Now we paste q' and q'' to define the desired map q :

$$q(tg_0H_\lambda + (1 - t)g_1H_\mu) = \begin{cases} q'((1 - 2t)g_0H_\lambda + 2tg_1H_\mu), & \text{if } 0 \leq t \leq 1/2, \\ q''((2t - 1)g_1H_\mu + (2 - 2t)g_0H_\lambda), & \text{if } 1/2 \leq t \leq 1. \end{cases}$$

Since $(1 - 2t)g_0H_\lambda + 2tg_1H_\mu$ and $(2t - 1)g_1H_\mu + (2 - 2t)g_0H_\lambda$ depend continuously upon $tg_0H_\lambda + (1 - t)g_1H_\mu \in \Delta$ and $q'(g_1A_\mu) = q''(g_0A_\lambda)$, we see that the continuity of q' and q'' implies the continuity of q . The equivariance of q is evident. Let us check that $q(y) \in \mathcal{P}(n)$.

Indeed, since for each $t \in [0, 1/2]$,

$$g_0A_\lambda \subset q(tg_0H_\lambda + (1 - t)g_1H_\mu) \subset B^n,$$

we infer that $q(tg_0H_\lambda + (1 - t)g_1H_\mu)$ belongs to $L_0(n)$. Further, for each $t \in [0, 1/2]$, the contact set of $q(tg_0H_\lambda + (1 - t)g_1H_\mu)$ is a subset of the contact set of g_0A_λ , and hence, it has an empty interior in the sphere S^{n-1} .

Analogously, for $t \in [1/2, 1]$,

$$g_1A_\mu \subset q(tg_0H_\lambda + (1 - t)g_1H_\mu) \subset B^n,$$

which implies that $q(tg_0H_\lambda + (1 - t)g_1H_\mu)$ belongs to $L_0(n)$. The contact set of $q(tg_0H_\lambda + (1 - t)g_1H_\mu)$ is a subset of the contact set of g_1A_μ , and hence, it also has an empty interior in the sphere S^{n-1} .

Thus, we have proved that for arbitrary $y \in \Gamma$, $q(y)$ belongs to $\mathcal{P}(n)$.

By Lemma 4.4, there is a G -map $R : \mathcal{N}(\mathcal{U}) \rightarrow \mathcal{F}_\infty(\Gamma)$ satisfying conditions (1) and (2) of Lemma 4.4. Observe that for every $y \in \mathcal{N}(\mathcal{U})$, $R(y)$ is a finite collection of elements of Γ . Now we define $\Phi' : \mathcal{N}(\mathcal{U}) \rightarrow \exp \mathbb{R}^n$ by

$$\Phi'(y) = \bigcup q(R(y)) = \bigcup_{x \in R(y)} q(x), \quad y \in \mathcal{N}(\mathcal{U}).$$

Then Φ' is well-defined, and by [23, Corollary 5.3.7] it is continuous. The equivariance of Φ' follows from that of R and of q .

Let $\Phi(y)$ be the convex hull of the set $\Phi'(y)$. Since $\Phi'(y) \subset \Phi(y) \subset B^n$, we infer that $\Phi(y) \in L_0(n)$.

The continuity of Φ now follows from the continuity of Φ' and of the convex hull operator (see [31, Theorem 2.7.4(iv)]).

Next, since $\Phi'(y)$ is a finite union of bodies from $\mathcal{P}(n)$, its contact set, i.e., the intersection $\partial B^n \cap \partial \Phi'(y)$, is the finite union of the contact sets of these bodies from $\mathcal{P}(n)$. Therefore, $\partial B^n \cap \partial \Phi'(n)$ has an empty interior in ∂B^n . It remains to observe that $\Phi(y)$ and $\Phi'(y)$ have the same contact set, and hence, $\Phi(y) \in \mathcal{P}(n)$.

Let us check that $\Phi(St(gH_\lambda, \mathcal{N}(\mathcal{U}))) \subset V(g, \lambda)$ for all $g \in G$, $\lambda \in \Lambda$. Here we need the following:

Claim. Let $O(X, a)$, $a > 0$, be the open a -ball in $\mathcal{B}(n)$ centered at $X \in \mathcal{B}(n)$. If $y = tg_0H_\lambda + (1-t)g_1H_\mu$ is a point of the 1-cell $\langle g_0H_\lambda, g_1H_\mu \rangle$ such that the corresponding sets g_0A_λ and g_1A_μ belong to $O(X, a)$, then $q(y) \in O(X, a) \cap L_0(n)$.

Proof of the Claim. First of all we observe that $O(X, a)$ is always a convex set in $\mathcal{B}(n)$, i.e., if $Y, Z \in O(X, a)$, then for every $t \in [0, 1]$ the convex body $tX + (1-t)Z$ belongs to $O(X, a)$.

Hence, $s_1(y) = tg_0A_\lambda + (1-t)A_\mu \in O(X, a)$, due to the convexity of $O(X, a)$. One also has $s_2(y) = g_0A_\lambda \in O(X, a)$.

On the other hand, according to ([31, Theorem 2.7.4(iv)]), the convex hull operator is non-expansive. So for $0 \leq t \leq 1/2$ we have

$$\begin{aligned} d_H(q'((1-2t)g_0H_\lambda + 2tg_1H_\mu), X) \\ \leq d_H(((1-2t)g_0A_\lambda + 2tg_1A_\mu) \cup g_0A_\lambda, X) < a. \end{aligned}$$

Similarly, for $1/2 \leq t \leq 1$ we have

$$\begin{aligned} d_H(q''((2t-1)g_1H_\mu + (2-2t)g_0H_\lambda), X) \\ \leq d_H(((2t-1)g_1H_\mu + (2-2t)g_0H_\lambda) \cup g_1A_\mu, X) < a. \end{aligned}$$

Therefore, $q(y)$ belongs to $O(X, a)$. Since $q(y) \in L_0(n)$, we infer finally that $q(y) \in O(X, a) \cap L_0(n)$, which completes the proof of the claim.

Now assume that $y \in St(gH_\lambda, \mathcal{N}(\mathcal{U}))$ is an arbitrary point. We have to show that $\Phi(y) \in V(g, \lambda)$.

Let

$$\tau = \langle g_0H_{\lambda_0}, \dots, g_nH_{\lambda_n} \rangle$$

be the carrier of y , where $g_0 = g$ and $\lambda_0 = \lambda$. Since $g_0S_{\lambda_0} \cap \dots \cap g_nS_{\lambda_n} \neq \emptyset$, it follows from (5.1) that

$$(5.2) \quad g_iX_{\lambda_i} \in St(g_0S_{\lambda_0}, \mathcal{U}) \subset V(g_0, \lambda_0)/2 \quad \text{for all } 0 \leq i \leq n.$$

Let $V(g_0, \lambda_0)$ be the intersection $L_0(n) \cap O(Y, \varepsilon/2)$ for some $Y \in L_0(n)$. It follows from (5.2) that

$$(5.3) \quad d_H(g_{\lambda_i} X_{\lambda_i}, Y) < \varepsilon/4 \quad \text{for all } 0 \leq i \leq n.$$

Since

$$d_H(g_i A_{\lambda_i}, g_i X_{\lambda_i}) = d_H(A_{\lambda_i}, X_{\lambda_i}) < \varepsilon/4,$$

it follows from (5.3) that

$$(5.4) \quad g_i A_{\lambda_i} \in O(Y, \varepsilon/2) \quad \text{for all } 0 \leq i \leq n.$$

By Lemma 4.4, $R(y)$ belongs to the 1-skeleton of the n -cell $\langle g_0 H_{\lambda_0}, \dots, g_n H_{\lambda_n} \rangle$, i.e., to the union

$$\bigcup_{i,j=0}^n \langle g_i H_{\lambda_i}, g_j H_{\lambda_j} \rangle.$$

Now, it follows from (5.4) and the above claim that the image under q of each 1-cell $\langle g_i H_{\lambda_i}, g_j H_{\lambda_j} \rangle$, $0 \leq i, j \leq n$, lies in $O(Y, \varepsilon/2)$. Since

$$\Phi'(y) = \bigcup_{x \in R(y)} q(x),$$

we infer that $\Phi'(y)$ also lies in $O(Y, \varepsilon/2)$. Since $\Phi(y)$ is the convex hull of $\Phi'(y)$ and Y is convex, the inequality $d_H(\text{conv } A, \text{conv } B) \leq d_H(A, B)$ (see [31, Theorem 2.7.4(iv)]) implies that $d_H(\Phi(y), Y) < \varepsilon/2$, i.e., $\Phi(y) \in O(Y, \varepsilon/2)$. Consequently,

$$\Phi(y) \in O(Y, \varepsilon/2) \cap L_0(n) = V(g_0, \lambda_0) = V(g, \lambda),$$

as required.

The second claim of Lemma 5.2 now follows immediately from the first one.

Indeed, for any $A \in L_0(n)$ there is $\lambda \in \Lambda$ such that $\varphi_\lambda(A) > 0$. Then $A \in G(S_\lambda)$, and hence, $A \in gS_\lambda$ for some $g \in G$. It then follows from the definition of the map $p : L_0(n) \rightarrow \mathcal{N}(\mathcal{U})$ that $p(A) \in St(gH_\lambda, \mathcal{N}(\mathcal{U}))$ (see the proof of Lemma 4.2). By the first statement, $\Phi(p(A)) \in V(g, \lambda)$. Since $A \in gS_\lambda \subset V(g, \lambda)$, we see that A and $\Phi(p(A))$ are ε -close. This means that the map Φp is ε -close to the identity map of $L_0(n)$, which completes the proof. \square

Let (X, d) be a metric space with a geodesic (or convex) metric d , i.e., for any two points $x, y \in X$ there is an isometry $\iota : [0, d(x, y)] \rightarrow X$ such that $\iota(0) = x$ and $\iota(d(x, y)) = y$. For any element $A \in \exp X$ the generalized closed r -ball centered at A is the set $A_r = \{x \in X \mid d(x, A) \leq r\}$. If $X = \mathbb{R}^n$ and $A \in \mathcal{B}(n)$, then A_r is just the parallel body $A + rB^n$, and hence, in this case, A_r is a compact, convex, centrally symmetric body.

Lemma 5.3. *Let (X, d) be a metric space with a geodesic (or convex) metric d . Then for any two elements $A, C \in \exp X$ and any two numbers $r, s > 0$, the following hold:*

- (1) $d_H(A_r, B_r) \leq d_H(A, B)$,
- (2) $d_H(A_r, A_s) \leq |r - s|$.

Proof. Since d is a convex metric, the first claim follows from [19, Proposition 10.5]. The second one follows from the property $(A_p)_q = A_{p+q}$ for any two nonnegative reals p and q (see [25, p. 38, Exercise 0.65.3(c)]). \square

Lemma 5.4. *For each $\varepsilon > 0$ there exist $O(n)$ -equivariant maps $f_\varepsilon, h_\varepsilon : L_0(n) \rightarrow L_0(n)$, ε -close to the identity map of $L_0(n)$, such that the images of f_ε and h_ε are disjoint.*

Proof. Define a continuous map $\gamma : L_0(n) \rightarrow \mathbb{R}$ by the rule

$$\gamma(A) = (1/2) \min\{d_H(B^n, A), \varepsilon\} \quad \text{for every } A \in L_0(n).$$

Let f_ε be just the closed $\gamma(A)$ -neighborhood of A in B^n , i.e.,

$$f_\varepsilon(A) = A_{\gamma(A)}.$$

By the choice of $\gamma(A)$, the set $f_\varepsilon(A)$ is different from B^n , and since $A \subset f_\varepsilon(A)$, we see that $f_\varepsilon(A) \in L_0(n)$. It is clear from the construction that f_ε is ε -close to the identity map of $L_0(n)$.

Let us check the continuity of f_ε . We have

$$d_H(f_\varepsilon(A), f_\varepsilon(C)) = d_H(A_{\gamma(A)}, C_{\gamma(C)}) \leq d_H(A_{\gamma(A)}, A_{\gamma(C)}) + d_H(A_{\gamma(C)}, C_{\gamma(C)}).$$

But by Lemma 5.3,

$$d_H(A_{\gamma(A)}, A_{\gamma(C)}) \leq |\gamma(A) - \gamma(C)| \quad \text{and} \quad d_H(A_{\gamma(C)}, C_{\gamma(C)}) \leq d_H(A, C).$$

Thus,

$$d_H(f_\varepsilon(A), f_\varepsilon(C)) \leq |\gamma(A) - \gamma(C)| + d_H(A, C).$$

Now the continuity of f_ε follows from that of γ . The $O(n)$ -equivariance of f_ε is immediate from the invariance of the metric d .

Next, we define the map $h_\varepsilon : L_0(n) \rightarrow L_0(n)$ to be the composition Φp from Lemma 5.2. Then $f_\varepsilon(A) \neq h_\varepsilon(C)$ for all $A, C \in L_0(n)$, since the contact set of $f_\varepsilon(A)$ has a nonempty interior in the boundary sphere $S^{n-1} = \partial B^n$ while the contact set of $h_\varepsilon(C)$ has an empty interior in S^{n-1} (this is because $h_\varepsilon(C) \in \mathcal{P}(n)$). \square

Lemma 5.5. *There is an $O(n)$ -equivariant strong deformation retraction (f_t) of $L(n)$ to its point B^n such that $f_t : L(n) \rightarrow L(n)$ is an $O(n)$ -isovariant map for all $0 < t \leq 1$. In particular, $f_t(A) \neq \{B^n\}$ for all $0 < t \leq 1$.*

Proof. For each $A \in L(n)$ and $0 \leq t \leq 1$, write

$$f_t(A) = (1-t)B^n + tA.$$

\square

Proof of Theorem 1.3. Since, by Theorem 2.4, $L(n)$ is an $O(n)$ -AR, we infer that it is also an H -AR (see, e.g., [30]). Therefore, by Theorem 2.7, $L(n)/H$ is an AR. Since $L_0(n)/H$ is an open subset of $L(n)/H$, it is a locally compact ANR. Now, in order to prove that $L_0(n)/H$ is a Q -manifold it suffices, according to Toruńczyk's Characterization Theorem [28], to check that for every $\varepsilon > 0$ there are continuous maps $f'_\varepsilon, h'_\varepsilon : L_0(n)/H \rightarrow L_0(n)/H$, ε -close to the identity map of $L_0(n)/H$, such that the images of f'_ε and h'_ε are disjoint. But this is immediate from Lemma 5.4, if we take for f'_ε and h'_ε the maps induced by f_ε and h_ε , respectively.

The $[0, 1)$ -stability of $L_0(n)/H$ follows from Lemma 5.5, which yields that the space $L_0(n)/H$ possesses an obvious proper deformation (preimage of each compact set is compact) to infinity:

$$(L_0(n)/H) \times [0, 1) \rightarrow L_0(n)/H.$$

Hence, by a result of R. Y. T. Wong [36], $L_0(n)/H$ is homeomorphic to its product with the half-open interval $[0, 1)$, i.e., $L_0(n)/H$ is $[0, 1)$ -stable. This completes the proof.

The following lemma for $n = 2$ was proved in [9]:

Lemma 5.6. *For each closed subgroup $K \subset O(n)$ acting non-transitively on S^{n-1} , and each $\varepsilon > 0$, there is a K -equivariant map $h_\varepsilon : L(n) \rightarrow L_0(n)$, ε -close to the identity map of $L(n)$. In particular, $h_\varepsilon(L(n)[K]) \subset L_0(n)[K]$.*

Proof. Let $r : \mathcal{B}(n) \rightarrow L(n)$ be the $O(n)$ -equivariant retraction from Theorem 2.3. Because $L(n)$ is compact (Theorem 2.4), one can find a real $0 < \delta < \varepsilon/2$ such that $d_H(r(A), A) < \varepsilon/2$ for all A belonging to the δ -neighborhood of $L(n)$ in $\mathcal{B}(n)$, where d_H denotes the Hausdorff metric on $\mathcal{B}(n)$.

Fix a centrally symmetric, convex polyhedron $P \subset \mathbb{R}^n$ with a nonempty interior, inscribing B^n , i.e., $P \subset B^n$ and all the vertices p_1, \dots, p_k of P lie on the unit sphere $S^{n-1} = \partial B^n$. Then the convex hull

$$R = \text{conv}(K(p_1) \cup \dots \cup K(p_k))$$

is a centrally symmetric, compact, convex, K -invariant subset of \mathbb{R}^n . Since it contains P , we see that R has a nonempty interior, and hence, $R \in \mathcal{B}(n)$. We claim that the boundary ∂R does not contain an $(n-1)$ -dimensional elliptic domain, i.e., an open connected subset of some $(n-1)$ -dimensional ellipsoid surface lying in \mathbb{R}^n . It suffices to show that none of the orbits $K(p_i)$, $i = 1, \dots, k$, contains an $(n-1)$ -dimensional elliptic domain. Assume the contrary, that $K(p_i)$ contains an $(n-1)$ -dimensional elliptic domain. Since $K(p_i)$ lies on the sphere S^{n-1} , then this domain should be in fact a domain of the sphere S^{n-1} . Since $K(p_i)$ is homogeneous and compact, we conclude that there are finitely many open subsets V_1, \dots, V_n of $K(p_i)$ such that $K(p_i) = V_1 \cup \dots \cup V_n$, where each V_j is homeomorphic to V . Next, by the Domain Invariance Theorem (see, e.g., [27, Ch. 4, Section 7, Theorem 16]), each V_j should be open in the sphere S^{n-1} , and hence, the union $V_1 \cup \dots \cup V_n$ is open in S^{n-1} . But $V_1 \cup \dots \cup V_n$ is also closed in S^{n-1} , because it is equal to $K(p_i)$. Now, by connectedness of S^{n-1} , it then follows that $K(p_i)$ is the whole sphere S^{n-1} . Consequently, K acts transitively on the unit sphere S^{n-1} , a contradiction. The claim is proved.

Now, let a be the distance of the origin from the boundary of R , and $T = (1/a)R$. Then T is a K -invariant, centrally symmetric, compact, convex body that circumscribes the unit ball B^n , i.e., $B^n \subset T$ and the boundaries ∂T and ∂B^n have a nonempty intersection.

Setting

$$h'(A) = A \cap (1 - \delta)T,$$

we obtain a map $h' : L(n) \rightarrow \mathcal{B}(n)$. Since T is a K -fixed point of $\mathcal{B}(n)$, we see that h' is K -equivariant.

Continuity of h' is evident.

Since $d_H(A, A \cap (1 - \delta)B^n) \leq \delta$ and $A \cap (1 - \delta)B^n \subset h'(A) \subset A$, we conclude that $d_H(A, h'(A)) \leq \delta$. In particular, h' is $(\varepsilon/2)$ -close to the inclusion $L(n) \hookrightarrow \mathcal{B}(n)$.

We claim that $h'(A)$ is not an ellipsoid for each $A \in L(n)$. Indeed, if $A \subset (1 - \delta)T$, then $A \neq B^n$, since T circumscribes B^n and $1 - \delta < 1$. On the other hand, $h'(A) = A$ in this case, and hence, $h'(A)$ is not an ellipsoid. If A is not contained in $(1 - \delta)T$, then the boundary of $h'(A)$ contains a domain lying in the boundary

of $(1 - \delta)T$. Since $(1 - \delta)T = ((1 - \delta)/a)R$ and since the boundary of R does not contain an $(n - 1)$ -dimensional elliptic domain (as shown above), we conclude that the boundary of $(1 - \delta)T$ does not contain an $(n - 1)$ -dimensional elliptic domain as well. Thus, $h'(A)$ is not an ellipsoid, and the claim is proved.

Since $r(h'(A))$ and $h'(A)$ have the same $GL(n)$ -orbit, we conclude that $r(h'(A)) \neq B^n$ for each $A \in L(n)$. Since r is $O(n)$ -equivariant and h' is K -equivariant, denoting by h_ε the composition rh' , we obtain a K -equivariant map $h_\varepsilon : L(n) \rightarrow L_0(n)$, ε -close to the identity map of $L(n)$. \square

Proof of Theorem 1.4. For the first claim it suffices to show that $L_0(n)/K$ is homeomorphic to Q_0 , the Hilbert cube with a removed point. By Theorem 1.3, $L_0(n)/K$ is a $[0, 1)$ -stable Q -manifold. On the other hand, Q_0 is a contractible $[0, 1)$ -stable Q -manifold. Therefore, according to a result of T. A. Chapman [12, Theorem 21.2], it remains only to check that $L_0(n)/K$ and Q_0 are homotopically equivalent, i.e., that $L_0(n)/K$ is contractible.

According to Theorem 2.4, $L(n)$ is an $O(n)$ -AR, which in turn implies that $L(n) \in K$ -AR (see, e.g., [30]). Then, by Theorem 2.7, $L(n)/K$ is an AR, and hence, is contractible. It follows from Lemma 5.6 that the singular point $\{B^n\}$ is a Z -set in the K -orbit space $L(n)/K$, and hence, according to [18], $L(n)/K$ and $L_0(n)/K$ have the same homotopy type. Since $L(n)/K$ is contractible, we see that $L_0(n)/K$ is contractible too.

For the second claim, it suffices to show that $L_0(n)[K]$ is homeomorphic to Q_0 . First we show that $L_0(n)[K]$ is a $[0, 1)$ -stable Q -manifold. Indeed, by Theorem 2.4, $L(n)$ is an $O(n)$ -AR, and hence, $L(n)[K]$ is an AR [2, Theorem 7]. It then follows that $L_0(n)[K]$ is a locally compact ANR. Now by Toruńczyk's Characterization Theorem [28], $L_0(n)[K]$ is a Q -manifold if we observe that the equivariant maps f_ε and h_ε from Lemma 5.4 take $L_0(n)[K]$ into itself. The $[0, 1)$ -stability of $L_0(n)[K]$ can be proved like that of $L_0(n)/K$ in Theorem 1.3. Indeed, Lemma 5.5 yields that the space $L_0(n)[K]$ possesses an obvious proper deformation to infinity

$$(L_0(n)[K]) \times [0, 1) \rightarrow L_0(n)[K].$$

Hence, by the result of R. Y. T. Wong [36], $L_0(n)/H$ is $[0, 1)$ -stable.

Let us show that $L_0(n)[K]$ is contractible. Since $L_0(n)[K] \neq \emptyset$, according to Lemma 5.6, the singular point $\{B^n\}$ is a Z -set in $L(n)[K]$. It then follows from [18] that $L(n)[K]$ and $L_0(n)[K]$ have the same homotopy type. But since $L(n)[K]$ is an AR, it is contractible, and hence, $L_0(n)[K]$ is contractible too. Since Q_0 also is a $[0, 1)$ -stable contractible Q -manifold, it only remains to apply the above-quoted result of T. A. Chapman to the Q -manifolds $L_0(n)[K]$ and Q_0 . This completes the proof.

6. PROOF OF THEOREM 1.6

We first prove the following fact.

Lemma 6.1. *$L_0(n)$ and $\Pi_0(n)$ have the same $O(n)$ -homotopy type.*

Proof. By [9, Lemma 4], there is an isovariant map $f : L_0(n) \rightarrow \Pi(n)$, yielding that the image of f lies, in fact, in $\Pi_0(n)$. Hence, the result follows from the following:

Claim. Every $O(n)$ -equivariant map $f : L_0(n) \rightarrow \Pi_0(n)$ is an $O(n)$ -homotopy equivalence.

To prove this claim we apply the James-Segal Theorem 2.5. In our case $G=O(n)$, $T=L_0(n)$ and $Z=\Pi_0(n)$.

By Theorem 2.4, $L(n) \in O(n)$ -AR, implying that $L_0(n) \in O(n)$ -ANR. Let $K \subset O(n)$ be a closed subgroup such that $L_0(n)[K] \neq \emptyset$. As we have seen above in the proof of Theorem 1.4, $L_0(n)[K]$ is contractible. Besides, it follows from the equivariance of f that $L_0(n)[K] \subset \Pi_0(n)[K]$, and so $\Pi_0(n)[K] \neq \emptyset$.

On the other hand, since $O(n)/H_i \in O(n)$ -ANR [26, p. 27], we infer that $\text{Cone}(O(n)/H_i) \in O(n)$ -AR (see [9, Lemma 3]). Consequently, $Q(H_i) \in O(n)$ -AR, $i \geq 1$, and hence, $\Pi(n) \in O(n)$ -AR, implying $\Pi_0(n) \in O(n)$ -ANR. Thus, it remains only to show that for each closed subgroup $K \subset O(n)$ with $\Pi_0(n)[K] \neq \emptyset$, $\Pi_0(n)[K]$ is contractible and, at the same time, $L_0(n)[K] \neq \emptyset$. We will show that in fact $\Pi(n)[K]$ is a Hilbert cube, implying the contractibility of $\Pi_0(n)[K]$. Indeed, it is not hard to see that if $\Pi_0(n)[K] \neq \emptyset$, then there is an orbit type (H_i) such that $(O(n)/H_i)[K] \neq \emptyset$. This implies that $(K) \preceq (H_i)$, and so, $K \subset gH_i g^{-1}$ for some $g \in O(n)$. But H_i occurs in $L_0(n)$ as a stabilizer. So there exists a body $A \in L_0(n)$ such that $O(n)_A = H_i$. Consequently, $K \subset O(n)_{gA}$. Since $gA \in L_0(n)$, we see that $L_0(n)[K] \neq \emptyset$. Next we have

$$\Pi(n)[K] = \prod_{i=1}^{\infty} (Q(H_i)[K]).$$

Since $O(n)/H_i \in O(n)$ -ANR, we see that $(O(n)/H_i)[K]$ is a nonempty ANR. Consequently,

$$\left(\text{Cone}(O(n)/H_i) \right) [K] = \text{Cone} \left((O(n)/H_i)[K] \right)$$

is a nondegenerate AR. Hence, according to a result of West [32], the countable product

$$Q(H_i)[K] = \left(\text{Cone}(O(n)/H_i)[K] \right)^{\infty}$$

is a Hilbert cube. This implies that $\prod_{i=1}^{\infty} (Q(H_i)[K])$ is a Hilbert cube, and hence, $\Pi_0(n)[K]$ is contractible. By applying the above-mentioned James-Segal Theorem, we complete the proof. \square

Proof of Theorem 1.6. Since by Theorem 1.3, $L_0(n)/H$ is a $[0, 1]$ -stable Q -manifold, according to Chapman's theorem [12, Theorem 21.2], it remains only to prove that $\Pi_0(n)/H$ is a $[0, 1]$ -stable Q -manifold of the same homotopy type as $L_0(n)/H$.

The fact that $\Pi_0(n)/H$ is a Q -manifold is proved in [9, Theorem A1]. Its $[0, 1]$ -stability follows from Wong's theorem [36] if we observe that $\Pi_0(n)/H$ possesses a proper deformation to infinity:

$$(\Pi_0(n)/H) \times [0, 1] \rightarrow \Pi_0(n)/H.$$

Indeed, this follows easily from the conic structure of $\Pi(n)$.

Finally, that $\Pi_0(n)/H$ and $L_0(n)/H$ have the same homotopy type follows immediately from Lemma 6.1.

7. PROOF OF THEOREM 1.9 AND REDUCTION OF CONJECTURE 1.8

We start with the following lemma.

Lemma 7.1. *Let G be a compact group, $N \subset G$ a closed, normal subgroup and X a G -ANR (resp., a G -AR). Then the N -fixed point set $X[N]$ is a G -ANR (resp., a G -AR) as well.*

Proof. According to [5, Corollary 5], there is a normed linear space L such that X can be embedded as a closed invariant subspace into $Z = C(G, L)$, the normed linear G -space of all continuous maps $f: G \rightarrow L$ endowed with the sup-norm and with the action gf of G defined by

$$(gf)(x) = f(xg); \quad f \in C(G, L) \quad g, x \in G.$$

Then there is a G -retraction $r: U \rightarrow X$ for some open G -neighborhood U of X in Z (resp., $U = Z$). Therefore, it suffices to prove that $Z[N]$ is a G -AR. One easily sees that $Z[N] = C(G/N, L)$, where the G -action $g\phi$ on $C(G/N, L)$ is defined by

$$(g\phi)(xN) = \phi(xgN), \quad \text{for } \phi \in C(G/N, L) \quad \text{and } g \in G, xN \in G/N.$$

Now $C(G/N, L)$ is a G -AR by [5, Theorem 8]. □

Let $\text{sexp } S^n$ be the subspace of $\exp S^n$ consisting of all centrally symmetric sets $A \subset S^n$, i.e., $A = -A$. By $\text{sexp}_0 S^n$ we will denote the complement $(\text{sexp } S^n) \setminus \{S^n\}$. Evidently, $\text{sexp } S^n$ is an $O(n)$ -invariant subset of $\exp S^n$.

The next lemma is immediate from Theorem 3.1 and Lemma 7.1 if we observe that $\text{sexp } S^{n-1} = (\exp S^{n-1})[N]$ with $N = \{1_{\mathbb{R}^n}, -1_{\mathbb{R}^n}\}$:

Lemma 7.2. *$\text{sexp } S^{n-1}$ is an $O(n)$ -AR.*

Lemma 7.3. *There exists an $O(n)$ -equivariant map $f: \text{sexp}_0 S^{n-1} \rightarrow L_0(n)$.*

Proof. For every $A \in \text{sexp}_0 S^{n-1}$, let

$$\varphi(A) = \text{conv}(A \cup B(0, 1/2)),$$

where $B(0, 1/2)$ is the closed $1/2$ -ball in \mathbb{R}^n centered at the origin, and conv stands for the convex hull. Clearly, φ is a well-defined, continuous, $O(n)$ -equivariant map of $\text{sexp}_0 S^{n-1}$ into $\mathcal{B}(n)$. Furthermore, $\varphi(A)$ is not an ellipsoid because $A \neq S^{n-1}$, implies that the boundary $\partial(\varphi(A))$ contains a nontrivial line segment. Now we set $f = r\varphi$, where $r: \mathcal{B}(n) \rightarrow L(n)$ is the $O(n)$ -equivariant map from Theorem 2.3. Since $\varphi(A)$ is not an ellipsoid, and since r preserves the $GL(n)$ -orbit, we conclude that $f(A)$ is not the unit ball B^n ; so $f(A) \in L_0(n)$. Since r and φ are $O(n)$ -equivariant, so is f . □

Lemma 7.4. *Let $K \subset O(n)$ be a closed subgroup. Then the following conditions are equivalent:*

- (1) K acts non-transitively on the sphere S^{n-1} ,
- (2) $(\exp_0 S^{n-1})[K] \neq \emptyset$,
- (3) $(\text{sexp}_0 S^{n-1})[K] \neq \emptyset$,
- (4) $L_0(n)[K] \neq \emptyset$.

Proof. (1) \implies (2). If K acts non-transitively, then there is a K -invariant proper subset $A \subset S^{n-1}$; so $A \in (\exp_0 S^{n-1})[K]$.

(2) \implies (3). If $A \in (\exp_0 S^{n-1})[K]$, then either

$$(-A) \cup A \neq S^{n-1} \quad \text{or} \quad (-A) \cap A \neq \emptyset,$$

and so at least one of the sets $(-A) \cup A$ and $(-A) \cap A$ belongs to $(\text{sexp}_0 S^{n-1})[K]$. Thus, $(\text{sexp}_0 S^{n-1})[K] \neq \emptyset$.

(3) \implies (4) is immediate from Lemma 7.3.

(4) \implies (1). If there is a body $A \in L_0(n)[K]$, then the contact set $A \cap S^{n-1}$ is a nonempty, K -invariant, proper subset of S^{n-1} . So the action of K on S^{n-1} is not transitive. \square

Lemma 7.5. *Every $O(n)$ -equivariant map $f : \text{sexp}_0 S^{n-1} \rightarrow L_0(n)$ is an $O(n)$ -homotopy equivalence.*

Proof. We are going to apply the James-Segal Theorem 2.5 with $G=O(n)$, $T = \text{sexp}_0 S^{n-1}$ and $Z = L_0(n)$. It follows from Lemma 7.2 that $\text{sexp}_0 S^{n-1} \in O(n)$ -ANR. Since $L(n) \in O(n)$ -AR (see Theorem 2.4), we see that also $L_0(n) \in O(n)$ -ANR.

Let $K \subset O(n)$ be a closed subgroup. Then by Lemma 7.4, $(\text{sexp}_0 S^{n-1})[K] \neq \emptyset$ iff $L_0(n)[K] \neq \emptyset$. Let $L_0(n)[K] \neq \emptyset$ and $K' = K \times \{1_{\mathbb{R}^n}, -1_{\mathbb{R}^n}\}$. Since

$$(\text{exp}_0 S^{n-1})[K'] = (\text{sexp}_0 S^{n-1})[K] \neq \emptyset,$$

the action of K' on S^{n-1} is not transitive, i.e., S^{n-1}/K' is not a singleton. Since $(\text{exp } S^{n-1})[K']$ is homeomorphic to $\text{exp}(S^{n-1}/K')$ and S^{n-1}/K' is a nondegenerate Peano continuum, by the Curtis-Schori-West Hyperspace Theorem (see, e.g., [23, §8.4]), $\text{exp}(S^{n-1}/K')$, and hence $(\text{sexp } S^{n-1})[K]$, is a Hilbert cube. Consequently, $(\text{sexp}_0 S^{n-1})[K]$, being a Hilbert cube with a removed point, is contractible.

According to Lemma 5.6, the singular point $\{B^n\}$ is a Z -set in $L(n)[K]$. Since $L(n)[K] \in \text{AR}$, it then follows from [18] that $L(n)[K]$ and $L_0(n)[K]$ have the same homotopy type, and hence, $L_0(n)[K]$ is contractible too. Now, by applying the above mentioned James-Segal Theorem 2.5, we complete the proof. \square

Similarly, the following can be proved:

Lemma 7.6. *Every $O(n)$ -equivariant map $f : \text{sexp}_0 S^{n-1} \rightarrow \text{exp}_0 S^{n-1}$ is an $O(n)$ -homotopy equivalence.*

Proof. By Lemma 7.4, for any closed subgroup $K \subset O(n)$ one has

$$(\text{sexp}_0 S^{n-1})[K] \neq \emptyset \iff (\text{exp}_0 S^{n-1})[K] \neq \emptyset.$$

As in the proof of Lemma 7.5, $(\text{sexp}_0 S^{n-1})[K]$, as well as $(\text{exp}_0 S^{n-1})[K]$, is contractible whenever $(\text{exp}_0 S^{n-1})[K] \neq \emptyset$ (or equivalently, $(\text{sexp}_0 S^{n-1})[K] \neq \emptyset$). Now apply the James-Segal Theorem 2.5. \square

Lemma 7.3 and Lemma 7.5 have the following immediate consequence.

Corollary 7.7. *$\text{sexp}_0 S^{n-1}$ and $L_0(n)$ have the same $O(n)$ -homotopy type.*

Analogously, Lemma 7.6 implies the following.

Corollary 7.8. *The natural inclusion $\text{sexp}_0 S^{n-1} \hookrightarrow \text{exp}_0 S^{n-1}$ is an $O(n)$ -homotopy equivalence.*

Proof of Theorem 1.9. Immediate from Corollaries 7.7 and 7.8.

Our final result reduces Conjecture 1.8 to an easier one:

Theorem 7.9. *For each closed subgroup $H \subset O(n)$, the two H -orbit spaces $L_0(n)/H$ and $(\text{exp}_0 S^{n-1})/H$ are homeomorphic iff $(\text{exp}_0 S^{n-1})/H$ is a Q -manifold.*

For the proof we need the following fact.

Lemma 7.10. *There exists an $O(n)$ -equivariant strong deformation retraction (f_t) of $\exp S^{n-1}$ to its $O(n)$ -fixed point $\{S^{n-1}\}$ such that $f_t(A) \neq \{S^{n-1}\}$ for all $0 < t \leq 1$ and $A \in \exp_0 S^{n-1}$.*

Proof. Observe that the usual spherical metric d on S^{n-1} is $O(n)$ -invariant and convex. For each $A \in \exp S^{n-1}$ and $0 \leq t \leq 1$, write

$$f_t(A) = \{x \in S^{n-1} \mid d(x, A) \leq (1-t)d_H(S^{n-1}, A)\}.$$

Due to the convexity of d this homotopy is continuous; it is also equivariant, since d and d_H are invariant. Other required properties of (f_t) are evident. \square

Proof of Theorem 7.9. Since by Theorem 1.3, $L_0(n)/H$ is a Q -manifold, only the “if” part requires a proof.

So, assume that $(\exp_0 S^{n-1})/H$ is a Q -manifold. It follows from Theorem 1.9 that

$$(\exp_0 S^{n-1})/H \quad \text{and} \quad L_0(n)/H$$

have the same homotopy type. Moreover, by Theorem 1.3, $L_0(n)/H$ is a $[0, 1)$ -stable Q -manifold. Therefore, according to Chapman’s theorem [12, Theorem 21.2], it only remains to see that $(\exp_0 S^{n-1})/H$ is $[0, 1)$ -stable too. Indeed, Lemma 7.10 yields that $(\exp_0 S^{n-1})/H$ possesses an obvious proper deformation to infinity:

$$((\exp_0 S^{n-1})/H) \times [0, 1) \rightarrow (\exp_0 S^{n-1})/H.$$

Now Wong’s result [36] implies that $(\exp_0 S^{n-1})/H$ is $[0, 1)$ -stable, and this completes the proof.

Acknowledgement. The author is grateful to the referee for helpful suggestions and remarks which led to a number of improvements of the text.

REFERENCES

1. H. Abels, *Parallelizability of proper actions, global K -slices and maximal compact subgroups*, Math. Ann. **212** (1974), 1–19. MR **51**:11460
2. S. A. Antonyan, *Retracts in categories of G -spaces*, Izvestiya Akad. Nauk Arm. SSR. Ser. Matem. **15** (1980), 365–378; English transl. in: Soviet J. Contemp. Math. Anal. **15** (1980), 30–43. MR **82f**:54027
3. S. A. Antonyan, *Equivariant generalization of Dugundji’s theorem*, Mat. Zametki **38** (1985), 608–616; English transl. in: Math. Notes **38** (1985), 844–848. MR **87a**:54053
4. S. A. Antonyan, *An equivariant theory of retracts*, in: Aspects of Topology (In Memory of Hugh Dowker), 251–269, London Math. Soc. Lecture Note Ser. **93**, Cambridge Univ. Press, Cambridge, 1985. MR **87e**:54090
5. S. A. Antonian, *Equivariant embeddings into G -AR’s*, Glasnik Matematički **22** (42) (1987), 503–533. MR **89k**:54041
6. S. A. Antonyan, *Retraction properties of an orbit space*, Matem. Sbornik **137** (1988), 300–318; English transl. in: Math. USSR Sbornik **65** (1990), 305–321. MR **89k**:54042
7. S. A. Antonyan, *Retraction properties of a space of orbits, II*, Russian Math. Surv. **48** (1993), 156–157. MR **95b**:54023
8. S. A. Antonyan, *The Banach-Mazur compacta are absolute retracts*, Bull. Acad. Polon. Sci. Ser. Math. **46** (1998), 113–119. MR **99d**:54020
9. S. A. Antonyan, *The topology of the Banach-Mazur compactum*, Fund. Math. **166**, no. 3 (2000), 209–232. MR **2001k**:57026
10. S. Banach, *Théorie des Opérations Linéaires*, Monografie Matematyczne, Warszawa, 1932. MR **97d**:01035 (reprint)
11. G. Bredon, *Introduction to compact transformation groups*, Academic Press, New York-London, 1972. MR **54**:1265

12. T. A. Chapman, *Lectures on Hilbert cube manifolds*, C. B. M. S. Regional Conference Series in Math., **28**, Amer. Math. Soc., Providence, RI, 1975. MR **54**:11336
13. D. W. Curtis, *Hyperspaces of noncompact metric spaces*, Compositio Math. **40** (1980), 139–152. MR **81c**:54009
14. D. W. Curtis, *Boundary sets in the Hilbert cube*, Topol. Appl. **20** (1985), 201–221. MR **87d**:57014
15. J. Dugundji, *Topology*, Allyn and Bacon Inc., Boston, 1966. MR **33**:1824
16. P. Fabel, *The Banach-Mazur compactum $Q(2)$ is an absolute retract*, in: Topology and Applications (International Topological Conference dedicated to P. S. Alexandroff's 100th birthday, Moscow, May 27–31, 1996), p. 57, Moscow, 1996.
17. R. E. Heisey and J. E. West, *Orbit spaces of the hyperspace of a graph which are Hilbert cubes*, Colloq. Math. **56** (1988), 59–69. MR **90a**:57024
18. D. W. Henderson, *Z-sets in ANR's*, Trans. Amer. Math. Soc. **213** (1975), 205–215. MR **52**:11830
19. A. Illanes and S. B. Nadler Jr., *Hyperspaces. Fundamentals and Recent Advances*, Marcel Dekker, Inc., New York-Basel, 1999. MR **99m**:54006
20. I. M. James and G. B. Segal *On equivariant homotopy theory*, Lecture Notes in Math. **788** (1980), 316–330. MR **82f**:55014
21. F. John, *Extremum problems with inequalities as subsidiary conditions*, in: F. John, Collected papers, **2** (ed. by J. Moser), 543–560, Birkhäuser, 1985. MR **10**:719b; MR **87f**:01107
22. T. Matumoto, *On G -CW complexes and a theorem of J.H.C. Whitehead*, J. Fac. Sci. Univ. Tokyo Sect. I A Math. **18** (1971), 363–374. MR **49**:9842
23. J. van Mill, *Infinite-dimensional topology. Prerequisites and Introduction*, North-Holland Publ. Co., Amsterdam-New York-Oxford-Tokyo, 1989. MR **90a**:57025
24. J. Milnor, *Construction of universal bundles, II*, Ann. Math. **63**(3) (1956), 430–436. MR **17**:1120a
25. S. B. Nadler, Jr., *Hyperspaces of sets*, Marcel Dekker, Inc., New York and Basel, 1978. MR **58**:18330
26. R. Palais, *The classification of G -spaces*, Memoirs of the Amer. Math. Soc. **36**, Providence, RI, 1960.
27. E. H. Spanier, *Algebraic Topology*, McGraw-Hill, New York, 1966. MR **35**:1007
28. H. Toruńczyk, *On CE -images of the Hilbert cube and characterization of Q -manifolds*, Fund. Math. **106** (1980), 31–40. MR **83g**:57006
29. H. Toruńczyk and J. E. West, *The fine structure of S^1/S^1 ; a Q -manifold hyperspace localization of the integers*, in: Proc. Internat. Conf. Geom. Topol., 439–449, PWN-Pol. Sci. Publ., Warszawa, 1980. MR **83g**:57005
30. J. de Vries, *Topics in the theory of topological transformation groups*, in: Topological Structures II, pp. 291–304, Math. Centre Tracts, Vol. **116**, Math. Centrum, Amsterdam, 1979. MR **81g**:54045
31. R. Webster *Convexity*, Oxford Univ. Press, Oxford, 1994. MR **98h**:52001
32. J. E. West, *Infinite products which are Hilbert cubes*, Trans. Amer. Math. Soc. **150** (1970), 1–25. MR **42**:1055
33. J. E. West, *Induced involutions on Hilbert cube hyperspaces*, Topology Proc. **1** (1976), 281–293. MR **58**:24276
34. J. E. West, *Open problems in infinite-dimensional topology*, in: Open Problems in Topology (ed. by J. van Mill and G. Reed), 524–586, North Holland, Amsterdam-New York-Oxford-Tokyo, 1990. MR **92c**:54001
35. M. Wojdyslawski, *Rétractes absolus et hyperespaces des continus*, Fund. Math. **32** (1939), 184–192.
36. R. Y. T. Wong, *Noncompact Hilbert cube manifolds*, preprint.

DEPARTAMENTO DE MATEMATICAS, FACULTAD DE CIENCIAS, UNAM, CIUDAD UNIVERSITARIA, MÉXICO D.F. 04510, MÉXICO

E-mail address: antonyan@servidor.unam.mx

LOCAL SOLVABILITY AND HYPOELLIPTICITY FOR SEMILINEAR ANISOTROPIC PARTIAL DIFFERENTIAL EQUATIONS

GIUSEPPE DE DONNO AND ALESSANDRO OLIARO

ABSTRACT. We propose a unified approach, based on methods from microlocal analysis, for characterizing the local solvability and hypoellipticity in C^∞ and Gevrey G^σ classes of 2-variable semilinear anisotropic partial differential operators with multiple characteristics. The conditions imposed on the lower-order terms of the linear part of the operator are optimal.

1. INTRODUCTION

We consider a class of semilinear anisotropic equations with multiple characteristics in two variables (x, y) belonging to the set $\Omega := \{\sqrt{x^2 + y^2} < \delta\}$, δ sufficiently small, of the form

$$(1.1) \quad P(x, y, D_x, D_y)u + F(x, y, \partial_x^l \partial_y^j u)|_{l\frac{m}{d}+j < m-t} = \mu f(x, y),$$

where the linear part is given by

$$(1.2) \quad P(x, y, D_x, D_y) = D_y^m - b_0(x, y)D_x^d + \sum_{m-t \leq l\frac{m}{d}+j < m} a_{lj}(x, y)D_x^l D_y^j,$$

with $m, d, j, l \in \mathbb{Z}_+$, $0 < t < \frac{1}{2}$, μ sufficiently small, $D_x = -i\frac{\partial}{\partial x}$, $D_y = -i\frac{\partial}{\partial y}$; we shall also say that $l\frac{m}{d} + j$ is the anisotropic order of $D_x^l D_y^j$, and so the nonlinearity involves derivatives of anisotropic order less than $m-t$. Our main aim is to propose a unified approach for a complete analysis of the influence of the lower-order terms of (1.2) on the solvability and hypoellipticity of (1.1) in the C^∞ category and in the Gevrey spaces G^σ beyond the critical index $m/(m-1)$. The arguments in our proofs are based mainly on microlocal tools: pseudo-differential operators, wave front sets, allowing relevant simplifications in the study, and $S_{\rho,\delta}^m$ techniques.

Some papers have been devoted to the study of this kind of problem; see, for example, Hounie-Santiago [HS] and Gramchev-Popivanov [GP] on the local solvability of semilinear partial differential equations in the case of simple characteristics, Gramchev-Rodino [GR] about Gevrey solvability for equations with multiple characteristics (see also Spagnolo [SP] and Kajitani-Spagnolo [KS]), and Garello [G] regarding the inhomogeneous elliptic case; see also Šananin [S] on the C^∞ local solvability of equations of quasi-principal type and Lorenz [LO] regarding anisotropic operators with characteristics of constant multiplicity.

Received by the editors February 7, 2001 and, in revised form, October 8, 2002.

2000 *Mathematics Subject Classification.* Primary 35S05.

We consider an F that is C^∞ and nonlinear, and we assume that the coefficients in (1.2) are C^∞ . We always suppose that

$$(1.3) \quad d < m, \quad \Re b_0(0,0) \neq 0, \quad F(x,y,0) = 0.$$

We recall that the nonzero requirement on b_0 is an invariant nondegeneracy condition, usually required in the study of the local solvability and hypoellipticity of the linear operator (1.2) in C^∞ and in the Gevrey classes G^σ , $\sigma > \frac{m}{m-1}$; see for example Liess-Rodino [LR2], De Donno-Rodino [DR2], in which Gevrey hypoellipticity for PDEs with high multiplicity is proved. Let us also observe that if $\Im b_0(x,y) \neq 0$, then the operator is quasi-elliptic; the results of hypoellipticity and local solvability are well known in this case. Regarding G^σ data, see for example Marcolongo-Oliaro [MO], in which the local solvability is proved in the n -dimensional case and under hypotheses on the quasi-principal symbol; in the present paper we admit less regular data $f(x,y)$ with respect to the case studied in [MO], but we add hypotheses on the lower-order terms. In this frame it will be convenient to use the Sobolev anisotropic space $H_{\frac{1}{q}}^s$, $q \geq 1$, defined by

$$\|f\|_{H_{\frac{1}{q}}^s} := \int (1 + |\xi|^{\frac{2}{q}} + |\eta|^2)^s |\widehat{f}(\xi, \eta)|^2 d\xi d\eta < \infty,$$

$\widehat{f}(\xi, \eta)$ being the Fourier transform of $f(x, y)$. For $s > \frac{1+q}{2}$, $H_{\frac{1}{q}}^s$ is an algebra; cf. the inhomogeneous Schauder estimates and Garello [G, Proposition 2.5]. Moreover, we define the anisotropic characteristic manifold

$$(1.4) \quad \Sigma_\delta := \{(x, y, \xi, \eta) \in \Omega \times \mathbb{R}^2 \setminus \{0\}, \eta^m - \Re b_0 \xi^d = 0\}.$$

We recall the definition of the Gevrey anisotropic space $G^{(q_1, q_2)}(\Omega)$.

Definition 1.1. Let $q_1 > 1$, $q_2 > 1$. We denote by $G^{(q_1, q_2)}(\Omega)$ the set of all the functions $f \in C^\infty(\Omega)$ such that the following condition holds: for every compact $K \subset \Omega$ there exists a positive constant C_K such that $\sup_K |\partial_x^l \partial_y^j f(x, y)| \leq C_K^{l+j+1} (l!)^{q_1} (j!)^{q_2}$ for every $l, j \in \mathbb{Z}_+$.

As usual, $G_0^{(q_1, q_2)}(\Omega)$ is the set of all the functions in $G^{(q_1, q_2)}(\Omega)$ with compact support in Ω .

Let us state the main results.

Theorem 1.1. Let $(l^*, j^*) \in \mathbb{Z}_+^2$ be the unique couple in \mathbb{Z}_+^2 having anisotropic order $\frac{k^*}{d} = l^* \frac{m}{d} + j^*$, with $d(m - \frac{1}{2}) < k^* = ml^* + dj^* < dm$. We assume that for $(x, y, \xi, \eta) \in \Sigma_\delta$:

$$(1.5) \quad \begin{aligned} & i) \Im a_{l^* j^*}(x, y) \neq 0, \\ & ii) \text{ for all } (l, j) \text{ such that } dj + ml > k^*, \\ & \Im a_{l^* j^*}(x, y) \Im a_{lj}(x, y) \xi^{l+l^*} \eta^{j+j^*} \geq 0, \\ & iii) \Im a_{l^* j^*}(x, y) \Im b_0(x, y) \xi^{d+l^*} \eta^{j^*} \leq 0. \end{aligned}$$

Choose and fix $t = m - \frac{k^*}{d}$ and $s > \frac{1}{2} \frac{d+m}{d} + \frac{k^*}{d}$. Then we can find $\delta_0 > 0$, depending on P and s , such that for every $f \in H_{\frac{d}{m}}^s(\Omega)$ with compact support in Ω the equation (1.1) admits a solution $u \in H_{\frac{d}{m}}^{s+m-r^*}$ with $r^* = m - \frac{k^*}{d}$, provided $\mu = 1$ if $F \equiv 0$ and $\mu \|f\|_{H_{\frac{d}{m}}^s} < \mu_0$ for some $0 < \mu_0 \ll 1$ depending on the nonlinear term F if $F \not\equiv 0$. Finally, the linear operator P in (1.2) is $C^\infty(\Omega)$ hypoelliptic, and if its coefficients are analytic, P is $G^\sigma(\Omega)$ hypoelliptic, $\sigma \geq \frac{d}{k^*-d(m-1)}$, i.e., if u is a distribution in Ω such that $Pu \in G^\sigma(\Omega)$, then $u \in G^\sigma(\Omega)$.

Remark 1.1. The conditions (1.5) in Theorem 1.1 could be illustrated in a simpler way, observing that actually one estimates the imaginary part of the symbol $\sum_{\frac{k^*}{d} \leq l \frac{m}{d} + j < m} a_{lj}(x, y) \xi^l \eta^j$ in the corresponding operator (1.2) on the quasi-conic characteristic set $\eta^m - \Re b_0 \xi^d = 0$, e.g., substituting $\xi = (\Re b_0)^{-\frac{1}{d}} \eta^{\frac{m}{d}}$, the condition (1.5) reads

$$(1.6) \quad \left| \Im \left(\sum_{\frac{k^*}{d} \leq l \frac{m}{d} + j < m} a_{lj} (\Re b_0)^{-\frac{1}{d}} \eta^{\frac{lm}{d}} \eta^j \right) \right| \geq C |\eta|^{\frac{k^*}{d}}, \quad \eta \gg 1,$$

with $k^* = ml^* + dj^*$. This clarifies, at least intuitively, the loss of derivatives $\frac{k^*}{d}$ in Theorem 1.1.

Remark 1.2. It is always possible to rephrase the previous assumptions in Theorem 1.1 directly on the coefficients of P . For example, if $\Re b_0 > 0$ and m, d are odd, the conditions *i*), *ii*), *iii*) are respectively equivalent to:

- i')* $\Im a_{l^*j^*}(x, y) > 0$ (< 0);
- ii')* for all (l, j) such that $dj + ml > k^* = dj^* + ml^*$, $\Im a_{lj}(x, y) \geq 0$ (≤ 0) for $j + j^*$ and $l + l^*$ both even or both odd, and $\Im a_{lj}(x, y) \equiv 0$ otherwise;
- iii')* $\Im b_0(x, y) \leq 0$ (≥ 0) for j^* and $d + l^*$ both even or both odd, and $\Im b_0(x, y) \equiv 0$ otherwise.

In the picture on the next page, which resembles the Newton polygon pictures, we show the geometrical meaning of hypothesis *ii*) in Theorem 1.1 (or equivalently assumption *ii')* in Remark 1.2). We consider the operator of order $m = 9$ with $d = 7$:

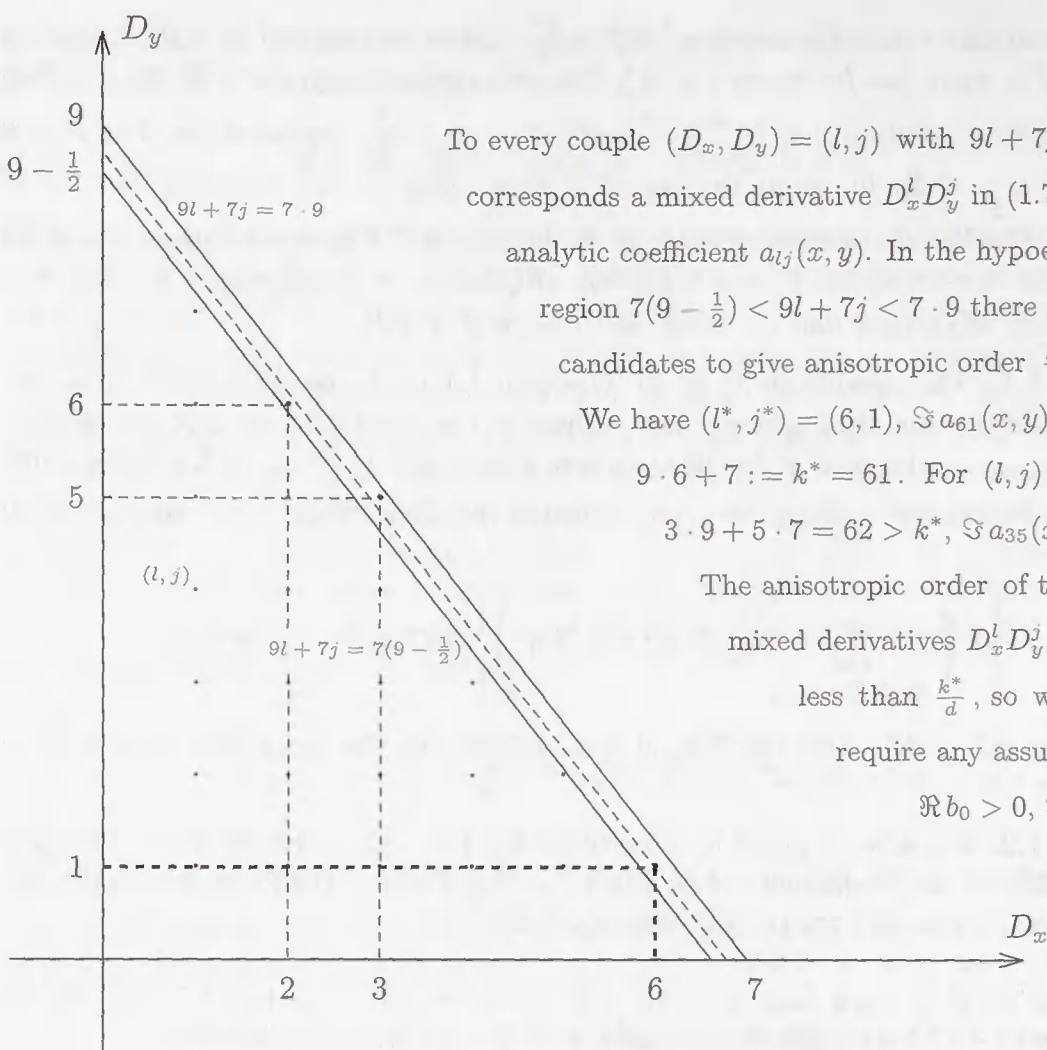
$$(1.7) \quad D_y^9 - (1 - iy^{2k}) D_x^7 + y^h D_x^3 D_y^5 + i D_x^6 D_y + \sum_{\frac{9}{7}l + j < \frac{61}{7}} a_{lj}(x, y) D_x^l D_y^j,$$

is C^∞ and Gevrey hypoelliptic and C^∞ solvable by Theorem 1.1.

We want to study now the case in which hypothesis *i*) in Theorem 1.1 is not satisfied: the basic idea is to refer to the Gevrey classes and transform the operator P in (1.2) into another operator that satisfies it. To this aim, we introduce the anisotropic Gevrey-Sobolev spaces $\mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))$, defined as the set of all L^2 functions for which

$$(1.8) \quad \|f\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}} := \|e^{\tau \psi(y, D_x)} f\|_{H_{\frac{1}{q}}^s} < +\infty,$$

where $q \geq 1$ is the Gevrey order, $s > 0$ the Sobolev index, and we take $r \in (0, 1)$, $\tau > 0$; $\psi = \psi(y, \xi)$ is a nonnegative function belonging to the Hörmander class $S_{1,0}^{\frac{\tau}{q}}((-\delta, \delta) \times \mathbb{R})$.



To every couple $(D_x, D_y) = (l, j)$ with $9l + 7j < 7 \cdot 9$, corresponds a mixed derivative $D_x^l D_y^j$ in (1.7) having analytic coefficient $a_{lj}(x, y)$. In the hypoellipticity region $7(9 - \frac{1}{2}) < 9l + 7j < 7 \cdot 9$ there are three candidates to give anisotropic order $\frac{k^*}{d} = \frac{k^*}{7}$. We have $(l^*, j^*) = (6, 1)$, $\Im a_{61}(x, y) > 0$ and $9 \cdot 6 + 7 = k^* = 61$. For $(l, j) = (3, 5)$, $3 \cdot 9 + 5 \cdot 7 = 62 > k^*$, $\Im a_{35}(x, y) \equiv 0$. The anisotropic order of the other mixed derivatives $D_x^l D_y^j$ is $\frac{9l+7j}{d}$, less than $\frac{k^*}{d}$, so we do not require any assumptions. $\Re b_0 > 0, \Im b_0 \leq 0$.

Theorem 1.2. *In the equation (1.1) let the datum f be in $G_0^{(\frac{m}{dr}, q_2)}(\Omega)$, $r \in (\frac{1}{2}, 1)$, $q_2 > 1$. Fix $t > 1 - r$ and assume that the coefficients of P in (1.2) are analytic and for $(x, y, \xi, \eta) \in \Sigma_\delta$ one of the following conditions holds:*

- [a] $\Im a_{lj}(x, y) \xi^l \eta^{j+m-1} \leq 0$ (≥ 0) for $dj + ml > (m - t)d$, and, moreover, $\Im b_0(x, y) \xi^d \eta^{m-1} \geq 0$ (≤ 0);
- [b] $\Im a_{lj}(x, y) (\text{sign } \xi) \xi^l \eta^{j+m-1} \leq 0$ (≥ 0) for $dj + ml > (m - t)d$, and $\Im b_0(x, y) (\text{sign } \xi) \xi^d \eta^{m-1} \geq 0$ (≤ 0).

Assume moreover that the nonlinear term F is analytic with respect to (x, y) , entire with respect to $\partial_x^l \partial_y^j u$, and (1.3) holds. Then, fixing ψ and taking s large, we can find $\delta_0 > 0$, depending on P and s , such that for every $f \in \mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}(\Omega)$ with compact support in Ω the semilinear equation (1.1) admits a solution $u \in \mathbb{H}_{\tau, \frac{m}{d}, r}^{s+m-(1-r), \psi}(\Omega)$ provided $\mu = 1$ if $F \equiv 0$, and $\mu \|f\|_{\mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}} < \mu_0$ for some $0 < \mu_0 \ll 1$ depending on the nonlinear term F if $F \not\equiv 0$.

The study of the weakly hyperbolic equation (1.1) resembles the study of degenerate parabolic equations, in that under suitable hypotheses equation (1.1) behaves like them in regard to local solvability and hypoellipticity in the C^∞ category and in

some Gevrey spaces G^σ , $\sigma > \frac{m}{m-2}$. See, for example, Gramchev-Popivanov-Yoshino [GPY, Section 3].

In Theorem 1.2 we require that the nonlinearity is analytic with respect to $(x, y, \partial_x^l \partial_y^j u)$. One encounters highly nontrivial difficulties in getting hard analysis type estimates on composition of nonanalytic Gevrey nonlinearities, and in the paper Gramchev-Rodino [GR] different Gevrey norms involving power series of finite Sobolev-type norms are used. A possible analogue of Theorem 1.2 could be proved in the anisotropic case for nonanalytic Gevrey nonlinearities, and this will be the subject of a future paper.

We point out that in hypotheses [a] and [b] of Theorem 1.2 the exponent $m-1$ plays the role of j^* in assumptions *ii*) and *iii*) of Theorem 1.1. In fact, for a suitable fixed $\psi(y, \xi) \in S_{1,0}^{\frac{r}{q}}((-\delta, \delta) \times \mathbb{R})$, the operator

$$\tilde{P} := e^{\tau\psi(y, D_x)} P(x, y, D_x, D_y) e^{-\tau\psi(y, D_x)}$$

contains all the terms of P and an additional pseudo-differential term

$$N_r(x, y) |D_x|^{\frac{d}{m}r} D_y^{m-1},$$

where $N_r(x, y)$ satisfies condition *i*) in Theorem 1.1.

Observe that for the existence of $l, j \in \mathbb{Z}_+$ such that $m-t \leq l\frac{m}{d} + j < m$, $t < \frac{1}{2}$, we have to require $m \geq 4$. The theorem applies, of course, also in the case when the assumption on the a_{lj} is empty; in this case we recapture the result of [MO].

Passing to the standard isotropic Gevrey classes $G^\sigma(\Omega)$, defined by the estimates

$$\sup_K |\partial_x^l \partial_y^j f(x, y)| \leq C^{l+j+1} (l!j!)^\sigma, \quad K \subset\subset \Omega,$$

we may conclude that equation (1.1) is locally solvable for $f \in G_0^{\frac{m}{d\tau}}(\Omega)$.

We observe that under assumption *i*) in Theorem 1.1 we obtain solvability in C^∞ , as well as C^∞ hypoellipticity, while for the operator \tilde{P} and Theorem 1.2 we only get local solvability in $G^{\frac{m}{d\tau}}$.

Remark 1.3. The meaning of conditions [a] and [b] in terms of the imaginary part of the coefficients of P depends on m, d , and the sign of $\Re b_0(0, 0)$. Let us analyze some situations.

- If m and d are even and $\Re b_0(0, 0) > 0$, we can treat equation (1.1) in one of the following cases:
 - (a) $\Im b_0 \equiv 0$, $\Im a_{lj} \leq 0$ (≥ 0) for l even and j odd, $\Im a_{lj} \equiv 0$ otherwise;
 - (b) $\Im b_0 \equiv 0$, $\Im a_{lj} \leq 0$ (≥ 0) for l and j odd, $\Im a_{lj} \equiv 0$ otherwise.
- If m is even, d is odd and $\Re b_0(0, 0) > 0$, since $p(x, y, \xi, \eta)$ is quasi-elliptic for $\xi < 0$, the hypotheses of Theorem 1.2 on the linear part become $\Im b_0 \equiv 0$, $\Im a_{lj} \leq 0$ (≥ 0) for j odd, $\Im a_{lj} \equiv 0$ for j even.

We also observe that if m and d are even and $\Re b_0(0, 0) < 0$, the operator is quasi-elliptic and we may apply known results; cf. Mascarello-Rodino [MR] and Rodino [RO].

Let us compare our result with the previously known ones. For the sake of brevity we limit our attention to a model of the form (1.2) with $d = m-1$, where

we fix attention on the case $x = 0, y = 0$:

(1.9)

$$D_y^m - D_x^{m-1} + i\alpha y^h D_x^{m-1} + i(\beta_1 y^q + \beta_2) D_y D_x^{m-2} + i\gamma D_y^2 D_x^{m-3} \\ + i(\mu_1 y^{2k} + \mu_2) D_y^3 D_x^{m-4}, \quad \alpha, \beta_1, \beta_2, \gamma, \mu_1, \mu_2 \in \mathbb{R} \quad (m \geq 3).$$

Let us analyze first the case $\beta_2 = \gamma = \mu_1 = \mu_2 = 0$. For $\alpha = \beta_1 = 0$ the operator (1.9) is not hypoelliptic; observe also that for $\alpha \neq 0, \beta_1 = 0, h = 1$, (1.9) is not hypoelliptic and not locally solvable, cf. Corli [C]. For $\alpha \neq 0, \beta_1 = 0, h$ even, (1.9) is hypoelliptic and locally solvable, despite the fact that $\Im b_0(0, 0) = 0$, cf. Menikoff [M], Popivanov [P1], Roberts [R]; for both $\alpha, \beta_1 \neq 0$ and h even, the operator (1.9) is not hypoelliptic if h is sufficiently large with respect to $q \geq 1$, cf. Popivanov-Popov [PP], Popivanov [P2], while for $h \leq \frac{m-1}{m-2}q$ it is hypoelliptic and locally solvable, cf. Gramchev-Popivanov [GP1].

Theorem 1.1 gives new conditions on the coefficients of the terms $D_y^j D_x^{m-j-1}$ for models of the type (1.2), to guarantee hypoellipticity; so now we discuss the case when $\beta_2, \gamma, \mu_1, \mu_2$ are not all zero.

Let us observe that, if j^* is odd, then (ii), (iii) in Theorem 1.1 actually imply $\Im a_{lj}(x, y) \equiv 0$ and $\Im b_0(x, y) \equiv 0$ for even $j < j^*$ ($d = m - 1$ implies $j + l = j^* + l^* = m - 1$); as examples of hypoelliptic and locally solvable operators characterized by Theorem 1.1, consider in this case (1.9) with $\alpha = \beta_1 = 0, \beta_2 \neq 0$ ($j^* = 1$), and (1.9) with $\alpha = \beta_2 = \gamma = \mu_1 = 0, \beta_1 \mu_2 > 0, q$ even ($j^* = 3$), having the same b_0 as the non-hypoelliptic operator $D_y^m - D_x^{m-1}$. If j^* is even, then (ii) implies $\Im a_{lj} \equiv 0$ for odd $j < j^*$; as a corresponding example of hypoelliptic and locally solvable operator consider (1.9) with $\beta_1 = \beta_2 = 0, \alpha \gamma > 0, h$ even ($j^* = 2$). The order m has to be chosen sufficiently large to satisfy the assumption $\frac{m-1}{2} > j^*$.

Now we discuss for the preceding examples the problem of local solvability in terms of Gevrey classes, arguing in the isotropic spaces $G^\sigma(\Omega)$. Concerning (1.9) with h even, to which we may add arbitrary perturbations of lower anisotropic order, we have σ -local solvability for $\sigma < \frac{m}{m-2}$. This follows from Theorem 1.2, and is also a consequence of Marcolongo-Oliaro [MO]. The result is sharp, in the sense that when $\alpha \neq 0$ and $h = 1$ (1.9) is not σ -locally solvable for $\sigma > \frac{m}{m-2}$, see Corli [C]. For $\alpha, \beta_1 \neq 0, \beta_2 = 0$, it was proved by Popivanov-Popov [PP] and Popivanov [P2] that (1.9) is not C^∞ locally solvable, and recently for h even, Marcolongo [MA] extended the result to σ -non-solvability for $\sigma > \frac{m}{m-2} + \varepsilon(h)$ with $\varepsilon(h) \rightarrow 0$ for $h \rightarrow \infty$. As new applications of our results in terms of G^σ locally solvable operators characterized by Theorem 1.2, consider (1.9) with $\beta_1 \neq 0, \alpha = \beta_2 = 0, q$ even ($m \geq 4$), which is σ locally solvable for $\sigma < \frac{m}{m-3}$, independently of lower anisotropic order perturbations. Compare in particular with $D_y^m - D_x^{m-1} + iy D_y D_x^{m-2}$, for which the change of sign of the imaginary part of the coefficient gives σ -non-solvability for $\sigma > \frac{m}{m-2}$. We have no change of sign and σ -solvability also in the interval $\frac{m}{m-3} > \sigma > \frac{m}{m-2}$. Moreover, let us consider (1.9) with $\beta_1 \mu_1 > 0, \alpha = \beta_2 = \mu_2 = 0, q$ even ($m \geq 10$), which is σ locally solvable for $\sigma < \frac{m}{m-5}$. The addition of lower-order anisotropic terms might produce non-solvability phenomena in C^∞ and G^σ for large σ . Observe finally that if the imaginary parts of a_{lj} in (1.2) vanish of high order at the origin, then the lower-order terms have no influence on the local solvability. As an example we consider the operator

$$(1.10) \quad D_y^5 - (1 - iy^k) D_x^4 + iy^l D_y D_x^3, \quad k \text{ even},$$

which is C^∞ locally solvable and C^∞ hypoelliptic for $k \leq \frac{4}{3}l$; see the arguments in the book by Gramchev and Popivanov [GP1, Theorem 3.1 and Chapter 4], and see also Theorem 3.1 in Gramchev [GG]. We also observe that by applying Theorem 1.2 we obtain Gevrey G^σ local solvability, $\sigma < \frac{5}{3}$, for $k > \frac{4}{3}l$, too.

2. GEVREY-SOBOLEV SPACES

As a preparation for the proof of Theorem 1.2, in this section we study a class of Gevrey-Sobolev spaces defined on the strip $\mathbb{R} \times (-\delta, \delta)$, $\delta > 0$. These spaces have been introduced in the n -dimensional case in [MO]; here we prove some results that will be used in the next sections for the local solvability of (1.1).

Definition 2.1. We define the Gevrey-Sobolev space $\mathbb{H}_{\tau,q,r}^{s,\psi}(\mathbb{R} \times (-\delta, \delta))$ as the set of all functions $f \in L^2(\mathbb{R} \times (-\delta, \delta))$ such that (1.8) holds. Writing $p = \frac{1}{q}$, we shall say that $\psi(y, \xi)$ is a weight function of order (r, p) . The operator $e^{\tau\psi(y, D_x)}$ acts on the function f in the following way:

$$e^{\tau\psi(y, D_x)} f(x, y) = \frac{1}{2\pi} \int e^{ix\xi} e^{\tau\psi(y, \xi)} \tilde{f}(y, \xi) d\xi,$$

where $\tilde{f}(y, \xi) = \int e^{-ix\xi} f(x, y) dx$, and $H_p^s(\mathbb{R} \times (-\delta, \delta))$, for s integer, is the space of all $g(x, y) \in L^2(\mathbb{R} \times (-\delta, \delta))$ such that

$$(2.1) \quad \|f(x, y)\|_{H_p^s}^2 := \sum_{k=0}^s \int (1 + |\xi|^{2p})^{s-k} |D_y^k \tilde{f}(y, \xi)|^2 d\xi dy < +\infty;$$

the definition of $H_p^s(\mathbb{R} \times (-\delta, \delta))$ extends to every $s > 0$ by interpolation.

Remark 2.1. The operator $e^{\tau\psi(y, D_x)}$ and its inverse $e^{-\tau\psi(y, D_x)}$ establish an isometry between the Hilbert spaces $\mathbb{H}_{\tau,q,r}^{s,\psi}(\mathbb{R} \times (-\delta, \delta))$ and $H_p^s(\mathbb{R} \times (-\delta, \delta))$.

We need to introduce suitable equivalent norms on the spaces $\mathbb{H}_{\tau,q,r}^{s,\psi}(\mathbb{R} \times (-\delta, \delta))$. First we recall that the following identities hold:

$$(2.2) \quad D_y^j e^{\tau\psi(y, D_x)} = \sum_{h=0}^j q_{j-h}^{(j)}(y, D_x) e^{\tau\psi(y, D_x)} D_y^h,$$

$$(2.3) \quad e^{\tau\psi(y, D_x)} D_y^j = \sum_{h'=0}^j r_{j-h'}^{(j)}(y, D_x) D_y^{h'} e^{\tau\psi(y, D_x)},$$

$$(2.4) \quad D_x^k e^{\tau\psi(y, D_x)} = e^{\tau\psi(y, D_x)} D_x^k$$

for every $k, j \in \mathbb{Z}_+$, where the symbols of the pseudo-differential operators $q_{j-l}^{(j)}(y, D_x)$ and $r_{j-l}^{(j)}(y, D_x)$ belong to the Hörmander class $S_{1,0}^{pr(j-l)}((-\delta, \delta) \times \mathbb{R})$. In particular, we obtain that

$$q_0^{(j)}(y, D_x) = 1, \quad q_1^{(j)}(y, D_x) = j\tau(D_y\psi)(y, D_x),$$

and

$$r_0^{(j)}(y, D_x) = 1, \quad r_1^{(j)}(y, D_x) = -j\tau(D_y\psi)(y, D_x).$$

For details, see [MO, Lemma 1.1], in which (2.2)-(2.4) are proved in the n -dimensional case.

Definition 2.2. We say that a function f belongs to the space $H^s_{(p_1,p_2)}(\mathbb{R}^2)$, $s \geq 0$, $0 < p_1 \leq 1$, $0 < p_2 \leq 1$, if and only if $f \in L^2(\mathbb{R}^2)$ and

(2.5)
$$\|f(x,y)\|^2_{H^s_{(p_1,p_2)}(\mathbb{R}^2)} := \int (1 + |\xi|^{2p_1} + |\eta|^{2p_2})^s |\widehat{f}(\xi,\eta)|^2 d\xi d\eta < \infty.$$

Lemma 2.1. Let $f \in H^s_{(p_1,p_2)}(\mathbb{R}^2)$, $s \geq 0$; let t and h be such that $\frac{t}{p_1} + \frac{h}{p_2} \leq s$. Then

$$\|\langle D_x \rangle^t \langle D_y \rangle^h f\|_{L^2(\mathbb{R}^2)} \leq \|\langle D_x \rangle^{p_1 s} f\|_{L^2(\mathbb{R}^2)} + \|\langle D_y \rangle^{p_2 s} f\|_{L^2(\mathbb{R}^2)} + \|f\|_{L^2(\mathbb{R}^2)},$$

where $\langle D_x \rangle^l$ and $\langle D_y \rangle^{l'}$ are the pseudo-differential operators with symbols $|\xi|^l$ and $|\eta|^{l'}$, respectively; we write (ξ, η) for the dual variables of (x, y) .

Proof. Let $K = [-1, 1] \times [-1, 1]$; we obtain

$$\begin{aligned} &\|\langle D_x \rangle^t \langle D_y \rangle^h f\|_{L^2(\mathbb{R}^2)}^2 \\ &\leq \int_{\mathbb{R}^2 \setminus K} (|\xi|^{2p_1 s} + |\eta|^{2p_2 s}) |\widehat{f}(\xi, \eta)|^2 d\xi d\eta + \int_K |\widehat{f}(\xi, \eta)|^2 d\xi d\eta \\ &\leq \|\langle D_x \rangle^{p_1 s} f\|_{L^2(\mathbb{R}^2)}^2 + \|\langle D_y \rangle^{p_2 s} f\|_{L^2(\mathbb{R}^2)}^2 + \|f\|_{L^2(\mathbb{R}^2)}^2. \end{aligned}$$

□

Remark 2.2. Let us suppose that $p_1, p_2 \in \mathbb{Q}$, and let s be a positive integer such that $p_j s$ is an integer, $j = 1, 2$. By Lemma 2.1 we have that an equivalent norm in $H^s_{(p_1,p_2)}(\mathbb{R}^2)$ is given by the following expression:

(2.6)
$$\|D_x^{p_1 s} f\|_{L^2(\mathbb{R}^2)} + \|D_y^{p_2 s} f\|_{L^2(\mathbb{R}^2)} + \|f\|_{L^2(\mathbb{R}^2)}.$$

Lemma 2.2. Consider $u \in H^s_p(\mathbb{R} \times (-\delta, \delta))$ with p a rational number and s a positive integer such that ps is an integer; we write Θ for $\mathbb{R} \times (-\delta, \delta)$. Then there exists a function $l_{s,\Theta} u \in H^s_{(p,1)}(\mathbb{R}^2)$ that extends u . Moreover, we can find a constant C such that:

$$\|l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)} \leq C \|u\|_{L^2(\mathbb{R} \times (-\delta, \delta))},$$

(2.7)
$$\|D_x^{ps} l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)} \leq C \|D_x^{ps} u\|_{L^2(\mathbb{R} \times (-\delta, \delta))},$$

$$\|D_y^s l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)} \leq C \|D_y^s u\|_{L^2(\mathbb{R} \times (-\delta, \delta))}.$$

This lemma can be proved using an argument similar to the one developed in the isotropic case by Egorov and Schulze in [ES, Theorem 27]. The proof is omitted.

Remark 2.3. Observe that $\|u\|_{L^2(\mathbb{R} \times (-\delta, \delta))} \leq \|l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)}$, $l_{s,\Theta} u$ being an extension of u ; so $\|u\|_{L^2(\mathbb{R} \times (-\delta, \delta))}$ and $\|l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)}$ are equivalent. In the same way we obtain that $\|D_x^{ps} u\|_{L^2(\mathbb{R} \times (-\delta, \delta))}$ and $\|D_y^s u\|_{L^2(\mathbb{R} \times (-\delta, \delta))}$ are equivalent to $\|D_x^{ps} l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)}$ and $\|D_y^s l_{s,\Theta} u\|_{L^2(\mathbb{R}^2)}$ respectively. So if s is an integer such that ps is an integer and $f \in H^s_p(\mathbb{R} \times (-\delta, \delta))$, using Lemma 2.2 and Remark 2.2 we can prove that an equivalent norm in the space $H^s_p(\mathbb{R} \times (-\delta, \delta))$ is given by the following expression:

(2.8)
$$\|D_x^{ps} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} + \|D_y^s f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} + \|f\|_{L^2(\mathbb{R} \times (-\delta, \delta))}.$$

Theorem 2.1. Let us fix $s > 0$, $\tau > 0$, $r \in (0, 1]$, $p \in (0, 1]$ and assume that s is a positive integer such that ps is an integer. Then the following norms are equivalent:

(a) $\|f\|_{\mathbb{H}^{s,\psi}_{\tau,q,r}(\mathbb{R} \times (-\delta, \delta))};$

- (b) $\sum_{\frac{l}{p}+j \leq s} \|D_x^l D_y^j e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))};$
 (c) $\sum_{\frac{l}{p}+j \leq s} \|e^{\tau\psi(y, D_x)} D_x^l D_y^j f\|_{L^2(\mathbb{R} \times (-\delta, \delta))};$
 (d) $\sum_{\frac{l_1+l_2}{p}+j_1+j_2 \leq s} \|D_x^{l_1} D_y^{j_1} e^{\tau\psi(y, D_x)} D_x^{l_2} D_y^{j_2} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))}.$

Proof. (a) \Leftrightarrow (b). Using Lemma 2.1 and Lemma 2.2 we obtain

$$\begin{aligned}
 & \sum_{\frac{l}{p}+j \leq s} \|D_x^l D_y^j e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} \\
 & \leq \sum_{\frac{l}{p}+j \leq s} \|D_x^l D_y^j l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} \\
 & \leq C (\|D_x^{ps} l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} + \|D_y^s l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} \\
 & \quad + \|l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)}) \\
 & \leq C_1 (\|D_x^{ps} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} + \|D_y^s e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} \\
 & \quad + \|e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))}) \\
 & \leq \|f\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))}.
 \end{aligned}$$

On the other hand, since $\|f\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))} \leq \|l_{s, \Theta}(e^{\tau\psi(y, D_x)} f)\|_{H_{(p, 1)}^s(\mathbb{R}^2)}$, by Remark 2.2 and Lemma 2.2 we have

$$\|f\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))} \leq C \sum_{\frac{l}{p}+j \leq s} \|D_x^l D_y^j e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))}.$$

(b) \Leftrightarrow (c). Using Lemma 2.1, Lemma 2.2, Remark 2.3 and the identities (2.2)-(2.4), we obtain

$$\begin{aligned}
 & \sum_{\frac{l}{p}+j \leq s} \|e^{\tau\psi(y, D_x)} D_x^l D_y^j f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} \\
 & \leq C \sum_{\frac{l}{p}+j \leq s} \sum_{h=0}^j \|r_{j-h}^{(j)}(y, D_x) D_y^h D_x^l e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))} \\
 & \leq C_1 \sum_{\frac{l}{p}+j \leq s} \sum_{h=0}^j \sum_{k=0}^{j-h} \|\langle D_x \rangle^{\tau p k + l} D_y^h l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} \\
 & \leq C_2 (\|D_x^{ps} l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} + \|D_y^s l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)} \\
 & \quad + \|l_{s, \Theta} e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R}^2)}) \\
 & \leq C_2 \sum_{\frac{l}{p}+j \leq s} \|D_x^l D_y^j e^{\tau\psi(y, D_x)} f\|_{L^2(\mathbb{R} \times (-\delta, \delta))}.
 \end{aligned}$$

In the opposite direction we may use similar arguments. By the same arguments we have (c) \Leftrightarrow (d). \square

Now we prove some important results that will be used in Section 3 for the solvability of the semilinear equation.

Theorem 2.2. *Let $0 < p < 1$ and $0 < r < 1$. Let $f \in G_0^{(\bar{q},q_2)}$ with $\bar{q} = \frac{1}{rp}$ and $q_2 > 1$. Then for all $s > 0$ and for every weight function $\psi(y, \xi)$ of order (r, p) there exists $\tau > 0$ such that $f \in \mathbb{H}_{\tau,q,r}^{s,\psi}$, where $q = \frac{1}{p}$.*

Proof. First we observe that $\mathbb{H}_{\tau,q,r}^{s,\psi} \subset \mathbb{H}_{\tau,q,r}^{t,\psi}$ for $s > t$; then, without loss of generality, we may assume that s is a positive integer such that ps is an integer. By Theorem 2.1 we can write

(2.9)
$$\|f\|_{\mathbb{H}_{\tau,q,r}^{s,\psi}} = \sum_{\frac{\alpha_1}{p} + \alpha_2 \leq s} \|e^{\tau\psi(y,\xi)} \widetilde{D^\alpha} f(y, \xi)\|_{L^2}.$$

The functions $D_{(x,y)}^\alpha f(x, y)$ obviously belong to $G_0^{(\bar{q},q_2)}(\mathbb{R} \times (-\delta, \delta))$ for every multi-index $\alpha = (\alpha_1, \alpha_2)$ such that $\frac{\alpha_1}{p} + \alpha_2 \leq s$, with the same constant C_K as in Definition 1.1, depending on s . For every integer k , applying the Fourier transform with respect to x we can find $b \in \mathbb{R}$, depending on $\text{supp } f$, such that

$$|\xi|^k |\widetilde{D^\alpha} f(y, \xi)| \leq \int_{-b}^b \overline{C}^{k+1} (k!)^{\bar{q}} dx \leq C (C^k (k!)^{\bar{q}}).$$

It follows immediately that $(1 + |\xi|)^k |\widetilde{D^\alpha} f(y, \xi)| \leq C (C_1^k (k!)^{\bar{q}})$ for all integers k , where we suppose $C \geq 1$. So we have

$$\sum_k \frac{1}{k!} \left(\frac{1 + |\xi|}{2C_1}\right)^{pkr} |\widetilde{D^\alpha} f(y, \xi)|^{rp} \leq C^{pr} \sum_k \left(\frac{1}{2^{rp}}\right)^k;$$

therefore we obtain

$$|\widetilde{D^\alpha} f(y, \xi)| \leq K e^{-M(1+|\xi|)^{rp}}.$$

Since $\psi(y, \xi)$ is a weight function of order (r, p) , we can find a constant C such that

$$|e^{\tau\psi(y,\xi)} \widetilde{D^\alpha} f(y, \xi)| \leq K e^{(\tau C - M)(1+|\xi|)^{rp}}.$$

Choosing $\tau < \frac{M}{2C}$, we can conclude that $e^{\tau\psi(y,\xi)} \widetilde{D^\alpha} f(y, \xi) \in L^2$ for $\frac{\alpha_1}{p} + \alpha_2 \leq s$; using (2.9), we have that $f \in \mathbb{H}_{\tau,q,r}^{s,\psi}$. □

Remark 2.4. Let $0 < p < 1$ and $0 < r < 1$. Let $f \in G_0^{(q_1,q_2)}(\mathbb{R} \times (-\delta, \delta))$ with $1 < q_1 < \frac{1}{rp}$ and $q_2 > 1$. Then, for every $s > 0$, for every weight function $\psi(y, \xi)$ of order (r, p) and for every $\tau > 0$, we have that $f \in \mathbb{H}_{\tau,q,r}^{s,\psi}(\mathbb{R} \times (-\delta, \delta))$, where $q = \frac{1}{p}$.

Theorem 2.3. *Let $\psi(y, \xi)$ be essentially subadditive with respect to ξ , i.e.,*

$$\psi(y, \xi_1 + \xi_2) \leq \psi(y, \xi_1) + \psi(y, \xi_2) + C,$$

cf. [GR]. Let $s_0 \geq \frac{p+1}{2p}$ satisfy the assumptions of Theorem 2.1. Then for every $s > s_0$ the space $\mathbb{H}_{\tau,q,r}^{s,\psi}(\mathbb{R} \times (-\delta, \delta))$ is an algebra, and there exists a constant C_s such that

(2.10)
$$\|uv\|_{\mathbb{H}_{\tau,q,r}^{s,\psi}} \leq C_s \|u\|_{\mathbb{H}_{\tau,q,r}^{s,\psi}} \|v\|_{\mathbb{H}_{\tau,q,r}^{s,\psi}}.$$

Proof. We begin by proving Theorem 2.3 with s a positive integer such that ps is an integer. Using Theorem 2.1, (c) and applying the Leibniz rule, we obtain

$$\|uv\|_{\mathbb{H}_{\tau,q,r}^{s,\psi}} \leq C \sum_{\frac{\beta_1+\gamma_1}{p} + \beta_2 + \gamma_2 \leq s} \|e^{\tau\psi(y,D_x)} (e^{-\tau\psi(y,D_x)} u_\beta e^{-\tau\psi(y,D_x)} v_\gamma)\|_{L^2},$$

where $u_\beta(x, y) = e^{\tau\psi(y, D_x)} D^\beta u$, $v_\gamma(x, y) = e^{\tau\psi(y, D_x)} D^\gamma v$ and the norms are in $\mathbb{R} \times (-\delta, \delta)$. Applying the Fourier transform with respect to x and writing $*_{(\xi)}$ for the convolution in the ξ variable, we have

$$\|uv\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}} \leq C_1 \sum_{\frac{\beta_1 + \gamma_1}{p} + \beta_2 + \gamma_2 \leq s} \|e^{\tau\psi(y, \xi)} ((e^{-\tau\psi(y, \xi)} \tilde{u}_\beta) *_{(\xi)} (e^{-\tau\psi(y, \xi)} \tilde{v}_\gamma))(y, \xi)\|_{L^2},$$

where \tilde{u}_β and \tilde{v}_γ stand for the partial Fourier transform with respect to x of u_β and v_γ respectively. The function $\psi(y, \xi)$ being essentially subadditive with respect to ξ , it follows immediately that $e^{\tau\psi(y, \xi) - \tau\psi(y, \xi - \mu) - \tau\psi(y, \mu)} \leq e^C$, and so

$$\|uv\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}} \leq C_2 \sum_{\frac{\beta_1 + \gamma_1}{p} + \beta_2 + \gamma_2 \leq s} \|u_\beta v_\gamma\|_{L^2}.$$

Since we have required $s > \frac{p+1}{2p}$, at least one of the inequalities $\frac{\beta_1}{p} + \beta_2 < s - \frac{p+1}{4p}$ and $\frac{\gamma_1}{p} + \gamma_2 < s - \frac{p+1}{4p}$ must be satisfied. Let $\rho > 0$ be such that

$$\begin{aligned} & \{(\alpha_1, \alpha_2) \in \mathbb{Z}_+^2 : \frac{\alpha_1}{p} + \alpha_2 < s - \frac{p+1}{4p}\} \\ &= \{(\alpha_1, \alpha_2) \in \mathbb{Z}_+^2 : \frac{\alpha_1}{p} + \alpha_2 \leq s - \frac{p+1}{4p} - \rho\}. \end{aligned}$$

Using Lemma 2.2 and Young's estimates, we have

$$\begin{aligned} \|uv\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}} &\leq C_3 \sum_{\substack{\frac{\beta_1}{p} + \beta_2 \leq s - \frac{p+1}{4p} - \rho \\ \frac{\gamma_1}{p} + \gamma_2 \leq s}} \|(l_{\frac{p+1}{4p} + \rho, \Theta} u_\beta)^\wedge\|_{L^1(\mathbb{R}^2)} \|(l_{0, \Theta} v_\gamma)^\wedge\|_{L^2(\mathbb{R}^2)} \\ &+ C_3 \sum_{\substack{\frac{\beta_1}{p} + \beta_2 \leq s \\ \frac{\gamma_1}{p} + \gamma_2 \leq s - \frac{p+1}{4p} - \rho}} \|(l_{0, \Theta} u_\beta)^\wedge\|_{L^2(\mathbb{R}^2)} \|(l_{\frac{p+1}{4p} + \rho, \Theta} v_\gamma)^\wedge\|_{L^1(\mathbb{R}^2)}. \end{aligned}$$

Since $u, v \in \mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))$ and $\frac{\beta_1}{p} + \beta_2 \leq s$, $\frac{\gamma_1}{p} + \gamma_2 \leq s$, Theorem 2.1 and (2.7) assure us that $\|(l_{0, \Theta} u_\beta)^\wedge\|_{L^2(\mathbb{R}^2)} \leq \|u\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}}$ and $\|(l_{0, \Theta} v_\gamma)^\wedge\|_{L^2(\mathbb{R}^2)} \leq \|v\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}}$. Moreover, if $\frac{\beta_1}{p} + \beta_2 \leq s - \frac{p+1}{4p} - \rho$, by the Hölder inequality and (2.7) we have

$$\begin{aligned} & \|(l_{\frac{p+1}{4p} + \rho, \Theta} u_\beta)^\wedge(\xi, \eta)\|_{L^1(\mathbb{R}^2)} \\ & \leq C_4 \|(1 + |\xi|^{p(\frac{p+1}{4p} + \rho)} + |\eta|^{p(\frac{p+1}{4p} + \rho)})(l_{\frac{p+1}{4p} + \rho, \Theta} u_\beta)^\wedge\|_{L^2(\mathbb{R}^2)} \leq C_5 \|u\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}}. \end{aligned}$$

The same arguments allow us to show that $\|(l_{\frac{p+1}{4p} + \rho, \Theta} v_\gamma)^\wedge\|_{L^1} \leq C \|v\|_{\mathbb{H}_{\tau, q, r}^{s, \psi}}$. So (2.10) holds. By interpolation, the result remains valid for every $s > s_0$. \square

3. ANALYSIS OF THE LINEAR EQUATION AND PROOF OF THEOREMS 1.1 AND 1.2

Let us consider the operator (1.2), where we take the coefficients $b_0(x, y)$ and $a_{lj}(x, y)$ in the space $G^{(q_1, q_2)}(\Omega)$ with $1 < q_1 < \frac{m}{dr}$, $q_2 > 1$. We choose t in (1.2) in such a way that there exist two integers \bar{l} and \bar{j} for which $\bar{l} \frac{m}{d} + \bar{j} = m - t$.

Let us observe that the operator $P(x, y, D_x, D_y)$ does not involve the terms $a_{lj}(x, y) D_x^l D_y^j$ with order $l \frac{m}{d} + j < m - t$; these terms have order $l \frac{m}{d} + j \leq m - t - \epsilon_t$,

where ϵ_t is given by the following expression:

$$(3.1) \quad \epsilon_t = \begin{cases} \min_{h \in I_t} \left(\mathcal{M} \left(\frac{d}{m} (m - t - h) \right)^{\frac{m}{d}} \right), & I_t \neq \emptyset, \\ 1, & I_t = \emptyset. \end{cases}$$

The symbol $\mathcal{M}(b)$ stands for the decimal part of b , and $I_t = \{h \in [0, m - t], h \in \mathbb{N} : \frac{d}{m}(m - t - h) \notin \mathbb{N}\}$. We deal with the local solvability at the origin of the equation

$$(3.2) \quad P(x, y, D_x, D_y)v(x, y) = f(x, y),$$

P as in (1.2); so it is not restrictive to multiply the coefficients $b_0(x, y)$ and $a_{lj}(x, y)$ by a function $\chi(x, y) \in G_0^{(q_1, q_2)}(\Omega)$ with support in a neighborhood of the origin. Thus, we can suppose that $b_0(x, y)$ and $a_{lj}(x, y)$ are compactly supported.

Now we fix a weight function $\psi(y, \xi)$ of order $(r, \frac{d}{m})$, essentially subadditive with respect to ξ ; for every $s \geq s_0$ and τ we consider the anisotropic Gevrey-Sobolev space $\mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))$.

For an arbitrary real number $q \geq 1$ we set:

- (i) $\mathbb{H}_{\tau, q, r, comp}^{s, \psi}(\Omega) := \{f \in \mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta)) \text{ with compact support contained in } \Omega\}$;
- (ii) $\mathbb{H}_{\tau, q, r, loc}^{s, \psi}(\Omega) := \{f \in \mathcal{D}'(\Omega) : \text{for every } \varphi \in G_0^{(q_1, q_2)}(\Omega), 1 < q_1 < \frac{q}{r}, q_2 > 1, \text{ we have } \varphi f \in \mathbb{H}_{\tau, q, r, comp}^{s, \psi}(\Omega)\}$;
- (iii) $\mathbb{H}_{\tau, q, r}^{s, \psi}(\Omega) := \{f \text{ such that } f \text{ is a restriction to } \Omega \text{ of a function belonging to } \mathbb{H}_{\tau, q, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta))\}$.

It follows from Theorem 2.1, Theorem 2.3 and Remark 2.4 that, for $s \geq s_0$,

$$(3.3) \quad P(x, y, D_x, D_y) : \mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta)) \longrightarrow \mathbb{H}_{\tau, \frac{m}{d}, r}^{s-m, \psi}(\mathbb{R} \times (-\delta, \delta)).$$

Moreover, the operator $P(x, y, D_x, D_y)$ can be regarded as a continuous map

$$P(x, y, D_x, D_y) : \mathbb{H}_{\tau, \frac{m}{d}, r, comp}^{s, \psi}(\Omega) \longrightarrow \mathbb{H}_{\tau, \frac{m}{d}, r, loc}^{s-m, \psi}(\Omega),$$

or also as a continuous map

$$P(x, y, D_x, D_y) : \mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}(\Omega) \longrightarrow \mathbb{H}_{\tau, \frac{m}{d}, r}^{s-m, \psi}(\Omega).$$

Now let us consider the following operator:

$$(3.4) \quad \tilde{P}(x, y, D_x, D_y) := e^{\tau\psi(y, D_x)} P(x, y, D_x, D_y) e^{-\tau\psi(y, D_x)}.$$

By Remark 2.1 and the previous considerations we have

$$P(x, y, D_x, D_y) = e^{-\tau\psi(y, D_x)} \tilde{P}(x, y, D_x, D_y) e^{\tau\psi(y, D_x)},$$

and, moreover,

$$(3.5) \quad \tilde{P}(x, y, D_x, D_y) : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^{s-m}(\mathbb{R} \times (-\delta, \delta)).$$

Now we want to analyze the conjugate operator $\tilde{P}(x, y, D_x, D_y)$; to this aim we start from the following proposition.

Proposition 3.1. *Let $0 < p \leq 1$ and $0 < r < 1$; as usual we set $q = \frac{1}{p}$. Let us fix a function $a(x, y) \in G_0^{(q_1, q_2)}$, $1 < q_1 < \frac{q}{r}$, $q_2 > 1$, and a weight function $\psi(y, \xi)$ of order (r, p) , essentially subadditive with respect to ξ . Then*

$$(3.6) \quad e^{\tau\psi(y, D_x)} a(x, y) e^{-\tau\psi(y, D_x)} = a(x, y) + Q_{-(q-r)}(x, y, D_x) + Q_{-2(q-r)}(x, y, D_x),$$

where the symbols $q_{-j(q-r)}(x, y, \xi)$ of the operators $Q_{-j(q-r)}(x, y, D_x)$, $j = 1, 2$, satisfy the following conditions:

$$q_{-(q-r)}(x, y, \xi) = \tau \partial_\xi \psi(y, \xi) (D_x a)(x, y) \in S_{1,0}^{-(q-r)p}(\Omega \times \mathbb{R});$$

$$q_{-2(q-r)}(x, y, \xi) \in S_{1,0}^{-2(q-r)p}(\Omega \times \mathbb{R}).$$

Proof. Setting $Q_a(x, y, D_x) = e^{\tau\psi(y, D_x)} a(x, y) e^{-\tau\psi(y, D_x)}$ and using the standard properties of the oscillatory integrals, we easily obtain that the symbol $q_a(x, y, \xi)$ of the operator $Q_a(x, y, D_x)$ is given by

$$(3.7) \quad q_a(x, y, \xi) = \int e^{ix\eta} e^{\tau\psi(y, \xi+\eta) - \tau\psi(y, \xi)} \tilde{a}(\eta, y) \bar{d}\eta,$$

where we write $\bar{d}\eta = (2\pi)^{-1} d\eta$ and we denote as usual by $\tilde{a}(\eta, y)$ the Fourier transform of $a(x, y)$ with respect to the x variable. Applying the Taylor formula to $e^{\tau\psi(y, \xi+\eta)}$, we obtain

$$(3.8) \quad e^{\tau\psi(y, \xi+\eta) - \tau\psi(y, \xi)} = 1 + \tau \partial_\xi \psi(y, \xi) \eta + \sum_{n=2}^{N-1} \frac{1}{n!} (e^{-\tau\psi(y, \xi)} \partial_\xi^n e^{\tau\psi(y, \xi)}) \eta^n + r_N(y, \xi, \eta),$$

where

$$\begin{aligned} & r_N(y, \xi, \eta) \\ &= \frac{\eta^N}{(N-1)!} \int_0^1 e^{-\tau\psi(y, \xi+t\eta)} \partial_\xi^N (e^{\tau\psi(y, \xi+t\eta)}) e^{\tau\psi(y, \xi+t\eta) - \tau\psi(y, \xi)} (1-t)^{N-1} dt. \end{aligned}$$

From (3.8) and the standard properties of the Fourier transform it follows immediately that

$$\begin{aligned} q_a(x, y, \xi) &= a(x, y) + \tau \partial_\xi \psi(y, \xi) (D_x a)(x, y) \\ &\quad + \sum_{n=2}^{N-1} \frac{1}{n!} \int e^{ix\eta} e^{-\tau\psi(y, \xi)} \partial_\xi^n e^{\tau\psi(y, \xi)} \widetilde{D_x^n a}(y, \eta) \bar{d}\eta \\ &\quad + \int e^{ix\eta} r_N(y, \xi, \eta) \tilde{a}(y, \eta) \bar{d}\eta \\ &= a(x, y) + q_{-(q-r)}(x, y, \xi) + (q_{-2(q-r)}^{(1)}(x, y, \xi) + q_{-2(q-r)}^{(2)}(x, y, \xi)). \end{aligned}$$

Using the Leibniz rule, Definition 1.1 and the fact that $\psi \in S_{1,0}^{rp}((-\delta, \delta) \times \mathbb{R})$, cf. Definition 2.1, we easily obtain that $q_{-(q-r)} \in S_{1,0}^{-(q-r)p}(\Omega \times \mathbb{R})$; so we have only to prove that $q_{-2(q-r)}^{(j)}(x, y, \xi) \in S_{1,0}^{-2(q-r)p}(\Omega \times \mathbb{R})$, $j = 1, 2$.

First we obtain by induction on n that for every $j, k \in \mathbb{Z}_+$ there exists a constant C_{jk} such that

$$(3.9) \quad |D_y^j D_\xi^k (e^{-\tau\psi(y, \xi)} \partial_\xi^n e^{\tau\psi(y, \xi)})| \leq C_{jk} (1 + |\xi|)^{prn - n - k}.$$

Using the Leibniz rule and the estimate (3.9), it is easy to deduce that for every fixed N , $q_{-2(q-r)}^{(1)}(x, y, \xi) \in S_{1,0}^{-2(q-r)p}(\Omega \times \mathbb{R})$.

Now we consider $q_{-2(q-r)}^{(2)}(x, y, \xi)$. Let us observe that, $\psi(y, \xi)$ being essentially subadditive with respect to ξ , we get $e^{\tau\psi(y, \xi+t\eta)-\tau\psi(y, \xi)} \leq e^{\overline{C}(1+|\eta|)^{pr}}$; moreover, due to $|D_y^j D_\xi^k \psi(y, \xi)| \leq C_{jk}(1+|\xi|)^{pr-k}$ ($\psi \in S_{1,0}^{rp}$), and to the inequality $\frac{1}{1+|\xi+t\eta|} \leq \frac{1+|\eta|}{1+|\xi|}$, we have that $|D_y^j D_\xi^k (\psi(y, \xi+t\eta) - \psi(y, \xi))| \leq \tilde{C}_{jk}(1+|\xi|)^{pr-k}(1+|\eta|)^{pr+k}$. Indeed,

$$\begin{aligned} |D_y^j D_\xi^k (\psi(y, \xi+t\eta) - \psi(y, \xi))| &\leq |D_y^j D_\xi^k (\psi(y, \xi+t\eta))| + |D_y^j D_\xi^k \psi(y, \xi)| \\ &\leq C_{jk}(1+|\xi+t\eta|)^{pr-k} + C_{jk}(1+|\xi|)^{pr-k} \\ &\leq 2^{pr} C_{jk}(1+|\xi|)^{pr}(1+|\eta|)^{pr} \left(\frac{1+|\eta|}{1+|\xi|} \right)^k + C_{jk}(1+|\xi|)^{pr-k} \\ &\leq \tilde{C}_{jk}(1+|\xi|)^{pr-k}(1+|\eta|)^{pr+k}. \end{aligned}$$

So we obtain, using Faà di Bruno's estimate, that

$$\begin{aligned} |D_y^j D_\xi^k (e^{\tau\psi(y, \xi+t\eta)-\tau\psi(y, \xi)})| &\leq \overline{C}_{jk} \sum_{0 < h \leq j+k} |e^{\tau\psi(y, \xi+t\eta)-\tau\psi(y, \xi)}| \\ &\times \sum_{\substack{j_1+\dots+j_h=j \\ k_1+\dots+k_h=k}} |D_y^{j_1} D_\xi^{k_1} (\psi(y, \xi+t\eta) - \psi(y, \xi))| \cdots |D_y^{j_h} D_\xi^{k_h} (\psi(y, \xi+t\eta) - \psi(y, \xi))| \\ &\leq \tilde{C}_{jk} e^{\overline{C}(1+|\eta|)^{pr}} (1+|\eta|)^{pr(j+k)+k} (1+|\xi|)^{pr(j+k)-k}. \end{aligned}$$

Using (3.9) with $\xi+t\eta$ instead of ξ , the previous estimate and the fact that $\frac{1}{1+|\xi+t\eta|} \leq \frac{1+|\eta|}{1+|\xi|}$, we have

$$(3.10) \quad \begin{aligned} |D_y^j D_\xi^k r_N(y, \xi, \eta)| \\ \leq C'_{jk} (1+|\eta|)^{pr(j+k)+prN+N+k} e^{\overline{C}(1+|\eta|)^{pr}} (1+|\xi|)^{pr(j+k)+prN-N-k}. \end{aligned}$$

Reasoning as in the proof of Theorem 2.2, we find that, if $a(x, y) \in G_0^{(q_1, q_2)}$, there exists a constant M such that $|D_y^j \tilde{a}(y, \eta)| \leq C_j e^{-M(1+|\eta|)^{p_1}}$, $p_1 = \frac{1}{q_1}$. Using this fact, the estimate (3.10) and the Leibniz rule, we have

$$\begin{aligned} |D_x^l D_y^j D_\xi^k q_{-2(q-r)}^{(2)}(x, y, \xi)| &\leq C'_{ljk} (1+|\xi|)^{pr(j+k)+prN-N-k} \\ &\times \int (1+|\eta|)^{pr(j+k)+prN+N+k} |\eta|^l e^{\overline{C}(1+|\eta|)^{pr}} e^{-M(1+|\eta|)^{p_1}} d\eta \\ &\leq C_{ljk} (1+|\xi|)^{pr(j+k)+prN-N-k}, \end{aligned}$$

since $pr < 1$, taking N sufficiently large, depending on j and k , we obtain that the symbol $q_{-2(q-r)}(x, y, \xi) = q_{-2(q-r)}^{(1)}(x, y, \xi) + q_{-2(q-r)}^{(2)}(x, y, \xi)$ is in the class $S_{1,0}^{-2(q-r)p}(\Omega \times \mathbb{R})$. \square

Definition 3.1. Let us consider a function $a(x, y, \xi, \eta) \in C^\infty(\Omega \times \mathbb{R}^2)$ and define

$$(3.11) \quad \lambda_p(\xi, \eta) = (1 + |\xi|^{2p} + |\eta|^2)^{\frac{1}{2}} \sim 1 + |\xi|^p + |\eta|.$$

We say that $a(x, y, \xi, \eta) \in S_{(p,1)}^{m,\mu}(\Omega \times \mathbb{R}^2)$, $p \leq 1$, if for every $l, j, k, h \in \mathbb{Z}_+$ there exists a constant C_{ljkh} such that

$$(3.12) \quad |D_x^l D_y^j D_\xi^k D_\eta^h a(x, y, \xi, \eta)| \leq C_{ljkh} \lambda_p(\xi, \eta)^m (1 + |\xi|)^{p\mu-k-ph}.$$

Proposition 3.2. *Let p, r , the function $a(x, y)$ and the weight function $\psi(y, \xi)$ be fixed as in Proposition 3.1. Then*

$$(3.13) \quad \begin{aligned} & e^{\tau\psi(y, D_x)} a(x, y) D_x^l D_y^j e^{-\tau\psi(y, D_x)} \\ &= a(x, y) D_x^l D_y^j - j\tau a(x, y) (D_y \psi)(y, D_x) D_x^l D_y^{j-1} \\ & \quad + A_{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(x, y, D_x, D_y), \end{aligned}$$

where the symbol $a_{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(x, y, \xi, \eta)$ of the pseudo-differential operator $A_{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(x, y, D_x, D_y)$ belongs to the class

$$S_{(p,1)}^{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(\Omega \times \mathbb{R}^2).$$

Remark 3.1. The operator $-j\tau a(x, y) (D_y \psi)(y, D_x) D_x^l D_y^{j-1}$ in (3.13) has the symbol $-j\tau a(x, y) (D_y \psi)(y, \xi) \xi^l \eta^{j-1}$ in the class $S_{(p,1)}^{\frac{l}{p}+j, -(1-r)}(\Omega \times \mathbb{R}^2)$, as is easy to prove since $\psi \in S_{1,0}^{rp}$.

Proof. Using the identities (2.2)-(2.4) and Proposition 3.1, we get

$$\begin{aligned} & e^{\tau\psi(y, D_x)} a(x, y) D_x^l D_y^j e^{-\tau\psi(y, D_x)} \\ &= e^{\tau\psi(y, D_x)} a(x, y) e^{-\tau\psi(y, D_x)} e^{\tau\psi(y, D_x)} D_x^l D_y^j e^{-\tau\psi(y, D_x)} \\ &= (a(x, y) + Q_{-(q-r)}(x, y, D_x) + Q_{-2(q-r)}(x, y, D_x)) \\ & \quad \times (D_x^l D_y^j - j\tau (D_y \psi)(y, D_x) D_x^l D_y^{j-1} + \sum_{h=0}^{j-2} q_{j-h}^{(j)}(y, D_x) D_x^l D_y^h) \\ &= a(x, y) D_x^l D_y^j - j\tau a(x, y) (D_y \psi)(y, D_x) D_x^l D_y^{j-1} \\ & \quad + A_{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(x, y, D_x, D_y). \end{aligned}$$

In the expression of the symbol of the last operator, let us analyze, for example, the term $a(x, y) q_{j-h}^{(j)}(y, \xi) \xi^l \eta^h$, for $h = 0, \dots, j-2$. We get, for $l', j', k' \in \mathbb{Z}_+$ and $h' \leq h$,

$$\begin{aligned} & |D_x^{l'} D_y^{j'} D_\xi^{k'} D_\eta^{h'} (a(x, y) q_{j-h}^{(j)}(y, \xi) \xi^l \eta^h)| \\ & \leq C_{l'j'k'h'} \sum_{j'_1+j'_2=j'} \sum_{\substack{k'_1+k'_2=k' \\ k'_2 \leq l}} |D_x^{l'} D_y^{j'_1} a(x, y)| |D_y^{j'_2} D_\xi^{k'_1} q_{j-h}^{(j)}(y, \xi)| |\xi|^{l-k'_2} |\eta|^{h-h'} \\ & \leq C_{l'j'k'h'} (1 + |\xi|)^{pr(j-h)-k'_1} (1 + |\xi|)^{l-k'_2} \lambda_p(\xi, \eta)^{h-h'} \\ & \leq C_{l'j'k'h'} \lambda_p(\xi, \eta)^{\frac{l}{p}+j-2(1-r)-h'} (1 + |\xi|)^{-k'} \\ & \leq C_{l'j'k'h'} \lambda_p(\xi, \eta)^{\frac{l}{p}+j} (1 + |\xi|)^{-2p(1-r)-k'-ph'}. \end{aligned}$$

In general, using the Leibniz rule, the estimates on $\psi \in S_{1,0}^{rp}$ and $q_{j-h} \in S_{1,0}^{pr(j-h)}$, and the fact that the symbols of the operators $q_{-k(q-r)}(x, y, D_x)$ are in the class $S_{1,0}^{-k(q-r)p}(\Omega \times \mathbb{R})$ for $k = 1, 2$, we obtain:

- $(a(x, y) + q_{-(q-r)}(x, y, \xi) + q_{-2(q-r)}(x, y, \xi)) \sum_{h=0}^{j-2} q_{j-h}^{(j)}(y, \xi) \xi^l \eta^h$ is in the class $S_{(p,1)}^{\frac{l}{p}+j, -2(1-r)}(\Omega \times \mathbb{R}^2)$;

- $(q_{-(q-r)}(x, y, \xi) + q_{-2(q-r)}(x, y, \xi))(\xi^l \eta^j - j\tau(D_y \psi)(y, \xi) \xi^l \eta^{j-1})$ is in the class $S_{(p,1)}^{\frac{l}{p}+j, -(q-r)}(\Omega \times \mathbb{R}^2)$.

Let us observe that $S_{(p,1)}^{\frac{l}{p}+j, -2(1-r)}(\Omega \times \mathbb{R}^2)$ and $S_{(p,1)}^{\frac{l}{p}+j, -(q-r)}(\Omega \times \mathbb{R}^2)$ are both contained in $S_{(p,1)}^{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(\Omega \times \mathbb{R}^2)$; so it follows immediately that

$$a_{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(x, y, \xi, \eta) \in S_{(p,1)}^{\frac{l}{p}+j, \max\{-2(1-r), -(q-r)\}}(\Omega \times \mathbb{R}^2). \quad \square$$

Now we want to analyze the behavior of the conjugate operator, defined by (3.4). We choose the weight function in the following way:

$$(3.14) \quad \psi(y, \xi) = \left(1 + \frac{y}{2\delta}\right) \varphi(\xi) |\xi|^{\frac{d}{m}r} \quad \left(\psi(y, \xi) = \left(1 - \frac{y}{2\delta}\right) \varphi(\xi) |\xi|^{\frac{d}{m}r}\right),$$

where $0 < r < 1$, $\delta > 0$ and $\varphi(\xi)$ is a C^∞ function such that $0 \leq \varphi(\xi) \leq 1$, $\varphi(\xi) = 0$ for $|\xi| \leq \frac{1}{2}$ and $\varphi(\xi) = 1$ for $|\xi| \geq 1$. Observe that $\psi(y, \xi)$ is a weight function of order $(r, \frac{d}{m})$, essentially subadditive with respect to ξ .

Proposition 3.3. *Let us fix the operator $P(x, y, D_x, D_y)$ as in (1.2) with $d < m$, and let us fix $0 < r < 1$; we choose the weight function as in (3.14). Then the symbol of the conjugate operator is given by*

$$\begin{aligned} \tilde{p}(x, y, \xi, \eta) &= p(x, y, \xi, \eta) - i \frac{m\tau}{2\delta} \varphi(\xi) |\xi|^{\frac{d}{m}r} \eta^{m-1} + p_{m, -(1-r)-\nu}(x, y, \xi, \eta) \\ (\tilde{p}(x, y, \xi, \eta) &= p(x, y, \xi, \eta) + i \frac{m\tau}{2\delta} \varphi(\xi) |\xi|^{\frac{d}{m}r} \eta^{m-1} + p_{m, -(1-r)-\nu}(x, y, \xi, \eta)), \end{aligned}$$

where $p(x, y, \xi, \eta)$ is the symbol of P and

$$p_{m, -(1-r)-\nu}(x, y, \xi, \eta) \in S_{(\frac{d}{m}, 1)}^{m, -(1-r)-\nu}(\Omega \times \mathbb{R}^2), \nu > 0.$$

Observe that

$$(i \frac{m\tau}{2\delta} \varphi(\xi) |\xi|^{\frac{d}{m}r} \eta^{m-1}) \in S_{(\frac{d}{m}, 1)}^{m, -(1-r)}(\Omega \times \mathbb{R}^2).$$

Proof. First we observe that we can write the operator $P(x, y, D_x, D_y)$ in the following way:

$$(3.15) \quad P(x, y, D_x, D_y) = D_y^m - b_0(x, y) D_x^d + \sum_{m-t \leq l \leq \frac{m}{d} + j \leq m - \epsilon_0} a_{lj}(x, y) D_x^l D_y^j,$$

where ϵ_0 is given by (3.1) with $t = 0$.

Now, applying Proposition 3.1 and Proposition 3.2 with $p = \frac{d}{m}$ and ψ as in (3.14), and using (2.4), we get

$$\begin{aligned} \tilde{P}(x, y, D_x, D_y) &= e^{\tau\psi(y, D_x)} D_y^m e^{-\tau\psi(y, D_x)} - e^{\tau\psi(y, D_x)} b_0(x, y) e^{-\tau\psi(y, D_x)} D_x^d \\ &+ \sum_{m-t \leq l \leq \frac{m}{d} + j \leq m - \epsilon_0} e^{\tau\psi(y, D_x)} a_{lj}(x, y) D_x^l D_y^j e^{-\tau\psi(y, D_x)} \\ &= D_y^m - m\tau(D_y \psi)(y, D_x) D_y^{m-1} + A_{m, \max\{-2(1-r), -(\frac{m}{d}-r)\}}(x, y, D_x, D_y) \\ &- b_0(x, y) D_x^d - Q_{-(\frac{m}{d}-r)}(x, y, D_x) D_x^d \\ &+ \sum_{m-t \leq l \leq \frac{m}{d} + j \leq m - \epsilon_0} a_{lj}(x, y) D_x^l D_y^j + A'_{m, -(1-r)-\epsilon_0}(x, y, D_x, D_y) \\ &= P(x, y, D_x, D_y) - m\tau(D_y \psi)(y, D_x) D_y^{m-1} + P_{m, -(1-r)-\nu}(x, y, D_x, D_y), \end{aligned}$$

where

$$\begin{aligned} a_{m, \max\{-2(1-r), -(\frac{m}{d}-r)\}}(x, y, \xi, \eta) &\in S_{(\frac{d}{m}, 1)}^{m, \max\{-2(1-r), -(\frac{m}{d}-r)\}}(\Omega \times \mathbb{R}^2), \\ q_{-(\frac{m}{d}-r)}(x, y, \xi) &\in S_{1,0}^{-(\frac{m}{d}-r), \frac{d}{m}}(\Omega \times \mathbb{R}), \\ a'_{m, -(1-r)-\epsilon_0}(x, y, \xi, \eta) &\in S_{(\frac{d}{m}, 1)}^{m, -(1-r)-\epsilon_0}(\Omega \times \mathbb{R}^2). \end{aligned}$$

By Definition 3.1, we can deduce that the symbol $p_{m, -(1-r)-\nu}(x, y, \xi, \eta) = a_{m, \max\{-2(1-r), -(\frac{m}{d}-r)\}}(x, y, \xi, \eta) - q_{-(\frac{m}{d}-r)}(x, y, \xi)\xi^d + a'_{m, -(1-r)-\epsilon_0}(x, y, \xi, \eta)$ is in the class $S_{(\frac{d}{m}, 1)}^{m, -(1-r)-\nu}(\Omega \times \mathbb{R}^2)$, where $\nu = \min\{1-r, \frac{m}{d}-1, \epsilon_0\}$. We observe that, since $r < 1$ and $d < m$, we have $\nu > 0$. \square

The following theorem allows us to find a parametrix of the operator (1.2) when $m - t = \frac{k^*}{d}$, for a positive integer k^* such that $d(m - \frac{1}{2}) < k^* < dm$ and the set $I_{k^*} := \{(l, j) \in \mathbb{Z}_+^2 : dj + ml = k^*\}$ consists of just one couple (l^*, j^*) . We take $b_0(x, y)$ and $a_{lj}(x, y)$ in $C^\infty(\Omega)$.

Theorem 3.1 (Sobolev parametrix). *Assume that (1.5) in Theorem 1.1 and (1.3) hold. Then there exists a linear map*

$$E_\infty : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^{s+\frac{k^*}{d}}(\mathbb{R} \times (-\delta, \delta))$$

such that

$$P(x, y, D_x, D_y)E_\infty u = \vartheta(x, y)u + R_\infty u,$$

where $\vartheta(x, y) \in C^\infty$, $\vartheta(x, y) = 1$ in a neighborhood of the origin and R_∞ is a regularizing map in the Sobolev anisotropic spaces.

Removing hypothesis *i*) in Theorem 1.1, we shall prove that there exists a Gevrey-Sobolev parametrix of the operator (1.2) for $t < \frac{1}{2}$, where now we take analytic coefficients, or more generally coefficients in the anisotropic Gevrey space $G_0^{(q_1, q_2)}(\Omega)$, $1 < q_1 < \frac{m}{dr}$, $q_2 > 1$ (cf. Proposition 3.1), with $r > \max\{\frac{1}{2}, 1-t-\epsilon_t\}$.

Theorem 3.2 (Gevrey-Sobolev parametrix). *Let one of the conditions [a] or [b] in Theorem 1.2 and (1.3) hold. Then there exists a linear map*

$$E : \mathbb{H}_{\tau, \frac{m}{d}, r}^{s, \psi}(\mathbb{R} \times (-\delta, \delta)) \longrightarrow \mathbb{H}_{\tau, \frac{m}{d}, r}^{s+m-(1-r), \psi}(\mathbb{R} \times (-\delta, \delta))$$

such that

$$P(x, y, D_x, D_y)Eu = \chi(x, y)u + Ru,$$

where $\chi(x, y) \in G_0^{(q_1, q_2)}(\Omega)$, $\chi(x, y) = 1$ in a neighborhood of the origin, and R is a regularizing map in the anisotropic Gevrey-Sobolev spaces.

The proof of Theorems 3.1 and 3.2 will be deduced from Theorem 3.3, below, about the hypoellipticity of the following class of differential polynomials with C^∞ coefficients $h_{(\cdot, \cdot)} : \Omega \rightarrow \mathbb{C}$:

$$(3.16) \quad p(x, y, \xi, \eta) = \eta^m - h_{d0}(x, y)\xi^d + \sum_{(l, j) \in I} h_{lj}(x, y)\xi^l \eta^j + \sigma(z, \zeta),$$

for $\zeta = (\xi, \eta)$ the dual variable of $z = (x, y)$. We define the following sets for $k \in \mathbb{R}_+$, $0 < k < dm$:

$$I_k = \{(l, j) \in \mathbb{R}_+ \times \mathbb{Z}_+ : dj + ml = k\},$$

and fix $k = k^*$ such that $d(m - \frac{1}{2}) < k^* < dm$. We use the notation k^- for all $k < k^*$ and k^+ for all $k > k^*$. We define $I = I_- \cup I_{k^*} \cup I_+$, with $I_- = \bigcup I_{k^-}$, $I_+ = \bigcup I_{k^+}$. The symbol $\sigma(x, y, \xi, \eta)$ in $C^\infty(\mathbb{R}^2 \times \mathbb{R}^2)$ is such that

$$(3.17) \quad |D_x^\alpha D_y^\beta D_\xi^\gamma D_\eta^\theta \sigma(z, \zeta)| \leq C_{\alpha\beta\gamma\theta} (1 + \lambda(\zeta))^{\overline{m} - (\gamma \frac{m}{d} + \theta)}, \quad \overline{m} < \frac{k^*}{d},$$

where $\lambda(\zeta) = |\xi|^{\frac{d}{m}} + |\eta|$ is the anisotropic norm; cf. the expression of the anisotropic Sobolev spaces in Definition 2.2: $H_{(\frac{d}{m}, 1)}^s(\mathbb{R}^2)$, $s \geq 0$. We recall that $\Sigma := \{(z, \zeta) \in \Omega \times \mathbb{R}^2 \setminus \{0\} : \eta^m - \Re b_{d0} \xi^d = 0\}$ is the anisotropic characteristic manifold of $p(x, y, \xi, \eta)$ in (3.16); letting Λ be a neighborhood of Σ , we denote by Γ the set $\Omega \times \Lambda$, and we state the following:

Theorem 3.3. *Assume that I_{k^*} consists of just one couple (l^*, j^*) , $k^* = dj^* + ml^*$, such that:*

$$i) \Im h_{l^*j^*}(x, y) \neq 0 \text{ for all } (x, y) \in \Omega,$$

$$ii) \text{ for all } (l, j) \text{ such that } dj + ml > k^* = dj^* + ml^*,$$

$$(3.18) \quad \Im h_{l^*j^*}(x, y) \Im h_{lj}(x, y) \eta^{j+j^*} \xi^{l+l^*} \geq 0, \quad (z, \zeta) \in \Gamma,$$

$$iii) \Im h_{l^*j^*}(x, y) \Im h_{d0}(x, y) \eta^{j^*} \xi^{d+l^*} \leq 0, \quad (z, \zeta) \in \Gamma,$$

$$iv) \Re h_{d0}(x, y) \neq 0, \text{ for all } (x, y) \in \Omega.$$

Then

$$(3.19) \quad |p(x, y, \xi, \eta)| \geq b \lambda(\zeta)^{\frac{k^*}{d}} \text{ in } \Omega \times \mathbb{R}^2,$$

for a suitable constant b . Also, for all $(\alpha, \beta) \in \mathbb{Z}_+^2$, $(\gamma, \theta) \in \mathbb{Z}_+^2$ and for all $K \subset\subset \Omega$ we have, with suitable constants $L_{\alpha, \beta, \gamma, \theta}$ and B , that

$$(3.20) \quad \frac{|D_x^\alpha D_y^\beta D_\xi^\gamma D_\eta^\theta p(x, y, \xi, \eta)| \lambda(\zeta)^{\rho(\gamma \frac{m}{d} + \theta) - \delta(\alpha \frac{m}{d} + \beta)}}{|p(x, y, \xi, \eta)|} \leq L_{\alpha, \beta, \gamma, \theta}, \quad |\xi| + |\eta| > B,$$

with $\rho = \frac{k^* - d(m-1)}{d}$, $\delta = \frac{dm - k^*}{d}$. Observe that $\delta < \rho$, since we have assumed $k^* > d(m - \frac{1}{2})$.

Remark 3.2. By formula (3.20) and by Mascarello and Rodino ([MR], Theorem 3.3.6), we have that the operator $P(z, D)$, associated to the symbol $p(z, \zeta)$ in (3.16), is C^∞ -microlocally hypoelliptic in Γ ; i.e., $\Gamma \cap WF Pu = \Gamma \cap WF u$, for all $u \in \mathcal{D}'(\Omega)$, where $WF u$ is the Hörmander wave front set. A microhypoelliptic operator is hypoelliptic too. If the coefficients are analytic and (3.17) holds for $C_{\alpha\beta\gamma\theta} = L^{\alpha+\beta+\gamma+\theta+1} \alpha! \beta! \gamma! \theta!$, we obtain G^σ -hypoellipticity, $\sigma \geq \frac{d}{k^* - d(m-1)}$.

Before starting the proof of Theorems 3.1 and 3.2, we also need the following class of symbols:

Definition 3.2. Let $a(x, y, \xi, \eta) \in C^\infty(\Omega \times \mathbb{R}^2)$. We say that $a(x, y, \xi, \eta) \in \mathcal{S}_{1,0}^{m; (p,1)}(\Omega \times \mathbb{R}^2)$ if for every $l, j, k, h \in \mathbb{Z}_+$ there exists a constant C_{ljkh} such that

$$(3.21) \quad |D_x^l D_y^j D_\xi^k D_\eta^h a(x, y, \xi, \eta)| \leq C_{ljkh} \lambda_p(\xi, \eta)^{m - \frac{k}{p} - h},$$

where $\lambda_p(\xi, \eta)$ is given by (3.11).

Further on we shall also refer to the microlocal classes of symbols $S_{(p,1)}^{m,\mu}(\Gamma)$, $S_{1,0}^{m;(p,1)}(\Gamma)$, where now Γ is a quasi-homogeneous cone. We leave to the reader the obvious definitions in this frame.

In the following we suppose that the quasi-homogeneous cone Γ is included in a region in which $|\xi|^p \geq c_0|\eta|$ (in particular we are interested in the case $p = \frac{d}{m}$); for these sets we have that $(1 + |\xi|)^p \sim \lambda_p(\xi, \eta)$, and so $S_{(p,1)}^{m,\mu}(\Gamma)$ can be identified with $S_{1,0}^{m+\mu;(p,1)}(\Gamma)$.

Definition 3.3. Let $q(x, y, \xi, \eta) \in S_{1,0}^{m;(p,1)}(\Gamma)$. Let us suppose that $m' \leq m$ and $0 \leq \delta < \rho \leq 1$. We say that $q(x, y, \xi, \eta)$ is of type (m, m', p, ρ, δ) if there exist positive constants $c, C, C_{ljk h}$ such that in Γ the following estimates hold:

$$(3.22) \quad |q(x, y, \xi, \eta)| \geq c \lambda_p(\xi, \eta)^{m'},$$

$$(3.23) \quad |D_x^l D_y^j D_\xi^k D_\eta^h q(x, y, \xi, \eta)| \leq C_{ljk h} |q(x, y, \xi, \eta)| \lambda_p(\xi, \eta)^{-\rho(\frac{k}{p}+h)+\delta(\frac{l}{p}+j)}$$

for $\lambda_p(\xi, \eta) \geq C$.

Proposition 3.4. Let $q(x, y, \xi, \eta)$ be of the type (m, m', p, ρ, δ) . Then there exists a symbol $q'(x, y, \xi, \eta) \in S_{\rho, \delta}^{-m';(p,1)}(\Gamma)$, i.e.,

$$|D_x^l D_y^j D_\xi^k D_\eta^h q'(x, y, \xi, \eta)| \leq C_{ljk h} \lambda_p(\xi, \eta)^{m-\rho(\frac{k}{p}+h)+\delta(\frac{l}{p}+j)},$$

q' being the parametrix of $q(x, y, \xi, \eta)$, i.e., $q \# q' \sim q' \# q \sim 1$, where $\#$ indicates the standard asymptotic product.

For the proof of this proposition, see Hunt and Piriou [HP].

Proof of Theorem 3.1 (Sobolev parametrix). Without loss of generality, we may assume $\rho^0 = (\xi_0, \eta_0)$ with $\xi_0 > 0$. The quasi-homogeneous conic neighborhood Γ of ρ^0 is then included in a region $\{\xi > 0, \eta^2 < C\xi^2\}$. Since P satisfies all the hypotheses of Theorem 3.3, we have that the symbol $p(x, y, \xi, \eta)$ is of the type $(m, \frac{k^*}{d}, \frac{d}{m}, \frac{k^*-d(m-1)}{d}, \frac{dm-k^*}{d})$. So using Proposition 3.4 we can find a linear map $E : H_{\frac{d}{m}}^s \longrightarrow H_{\frac{d}{m}}^{s+\frac{k^*}{d}}$ such that $PE = \vartheta(x, y) \varrho(D_x, D_y) + R$, where $\vartheta(x, y) \in C_0^\infty(\Omega)$, $\varrho(\xi, \eta) \in C^\infty(\mathbb{R}^2)$ with support in a quasi-homogeneous conic neighborhood of (ξ_0, η_0) , $R : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^t(\mathbb{R} \times (-\delta, \delta))$ for every $t \geq 0$. So we can construct two operators E_1 and R_1 such that $PE_1 = \vartheta(x, y) \varrho(D_x, D_y) + R_1$, where we can choose the function $\varrho(\xi, \eta) \in C^\infty$ quasi-homogeneous in (ξ, η) of order $(\frac{d}{m}, 1)$ and of degree 0 for large $(|\xi|^{\frac{d}{m}} + |\eta|)$, $\varrho(\xi, \eta) = 0$ in a quasi-homogeneous conic neighborhood of $\xi = 0$ and $R_1 : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^t(\mathbb{R} \times (-\delta, \delta))$ for every $t \geq 0$. We can suppose that $\text{supp}(1 - \varrho(\xi, \eta)) \subset \Gamma_0$, where Γ_0 is a sufficiently small neighborhood of $\xi = 0$ such that $\eta^m - b_0(x, y)\xi^d$ is quasi-elliptic in Γ_0 . In the following we denote by $\bar{p}_m(x, y, \xi, \eta)$ the anisotropic principal symbol of $P(x, y, D_x, D_y)$, i.e., $\bar{p}_m(x, y, \xi, \eta) = \eta^m - b_0(x, y)\xi^d$. By the results of Hunt and Piriou [HP] we have, in Γ_0 ,

$$\begin{aligned} p(x, y, \xi, \eta) \# \bar{p}_m^{-1}(x, y, \xi, \eta) \\ &= (\bar{p}_m(x, y, \xi, \eta) + q_{m-\epsilon_m}(x, y, \xi, \eta)) \# \bar{p}_m^{-1}(x, y, \xi, \eta) \\ &= 1 + q_{-\epsilon_m}(x, y, \xi, \eta), \end{aligned}$$

where $q_{-\epsilon_m} \in S_{0,0}^{-\epsilon_m;(\frac{d}{m},1)}$ and ϵ_m is given by (3.1). In a standard way we can construct $s_{-\epsilon_m} \in S_{0,0}^{-\epsilon_m;(\frac{d}{m},1)}$ such that

$$(1 + q_{-\epsilon_m}(x, y, \xi, \eta)) \# (1 + s_{-\epsilon_m}(x, y, \xi, \eta)) \sim 1.$$

Let us consider now the symbol

$$e_0 = \overline{p}_m^{-1}(x, y, \xi, \eta) \# (1 + s_{-\epsilon_m}(x, y, \xi, \eta)) \# \vartheta(x, y) (1 - \varrho(\xi, \eta))$$

and let E_0 be the pseudo-differential operator of symbol e_0 . By construction we obtain $PE_0 = \vartheta(x, y) (1 - \varrho(D_x, D_y)) + R_0$, with R_0 regularizing on the anisotropic Sobolev spaces $H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta))$. Taking $E_\infty = E_1 + E_0$, we have that

$$PE_\infty = PE_1 + PE_0 = \vartheta(x, y) + R_\infty,$$

where $R_\infty = R_1 + R_0$ is regularizing on $H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta))$. □

Proof of Theorem 3.2 (Gevrey-Sobolev parametrix). Let us suppose that one of the global conditions [a] or [b] holds. When [a] is satisfied we fix the weight function as in (3.14); if [b] holds we choose $\psi(y, \xi) = (1 + \frac{y}{2\delta} \operatorname{sign} \xi) \varphi(\xi) |\xi|^{\frac{d}{m}r}$ ($\psi(y, \xi) = (1 - \frac{y}{2\delta} \operatorname{sign} \xi) \varphi(\xi) |\xi|^{\frac{d}{m}r}$). By Proposition 3.3, the symbol $\tilde{p}(x, y, \xi, \eta)$ of the conjugate operator $\tilde{P}(x, y, D_x, D_y)$ defined by (3.4) satisfies all the hypotheses of Theorem 3.3 with $j^* = m - 1$, $l^* = \frac{d}{m}r$. So $\tilde{p}(x, y, \xi, \eta)$ is of type $(m, m - (1 - r), p, r, 1 - r)$. As in the first part of the proof of Theorem 3.1, by Proposition 3.4 we can find a linear map

$$\tilde{E} : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^{s+m-(1-r)}(\mathbb{R} \times (-\delta, \delta))$$

such that

$$\tilde{P}\tilde{E} = \chi(x, y) \kappa(D_x, D_y) + \tilde{R},$$

where $\chi(x, y)$ is arbitrarily fixed in $C_0^\infty(\Omega)$, $\kappa(\xi, \eta)$ is arbitrarily fixed in $C^\infty(\mathbb{R}^2)$ with support in a quasi-homogeneous conic neighborhood of (ξ_0, η_0) , and \tilde{R} is a regularizing operator in the anisotropic Sobolev spaces $H_{\frac{d}{m}}^s$. So we can find \tilde{E}_1 and \tilde{R}_1 such that $\tilde{P}\tilde{E}_1 = \chi(x, y) \kappa(D_x, D_y) + \tilde{R}_1$, where we can choose the C^∞ function $\kappa(\xi, \eta)$ with the properties of the function $\varrho(x, y)$ in the previous proof ($\kappa(\xi, \eta)$ quasi-homogeneous of degree 0 out of the origin, $\kappa(\xi, \eta) = 0$ in a neighborhood of $\xi = 0$ and $\operatorname{supp}(1 - \kappa(\xi, \eta)) \subset \Gamma_0$, where $\overline{p}_m(x, y, \xi, \eta)$ is quasi-elliptic in Γ_0); moreover, $\tilde{R}_1 : H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta)) \longrightarrow H_{\frac{d}{m}}^t(\mathbb{R} \times (-\delta, \delta))$ for every $t \geq 0$.

Reasoning as in the preceding proof and using Proposition 3.3 and the results of Hunt and Piriou [HP], we have, in Γ_0 ,

$$\begin{aligned} &\tilde{p}(x, y, \xi, \eta) \# \overline{p}_m^{-1}(x, y, \xi, \eta) \\ &= (\overline{p}_m(x, y, \xi, \eta) + q_{m-\min\{\epsilon_m, 1-r\}}(x, y, \xi, \eta)) \# \overline{p}_m^{-1}(x, y, \xi, \eta) \\ &= 1 + q_{-\min\{\epsilon_m, 1-r\}}(x, y, \xi, \eta), \end{aligned}$$

where the symbol $q_{-\min\{\epsilon_m, 1-r\}}$ is in the class $S_{0,0}^{-\min\{\epsilon_m, 1-r\};(\frac{d}{m},1)}$. In a standard way we can construct $s_{-\min\{\epsilon_m, 1-r\}} \in S_{0,0}^{-\min\{\epsilon_m, 1-r\};(\frac{d}{m},1)}$ such that

$$(1 + q_{-\min\{\epsilon_m, 1-r\}}(x, y, \xi, \eta)) \# (1 + s_{-\min\{\epsilon_m, 1-r\}}(x, y, \xi, \eta)) \sim 1.$$

Now let us consider the pseudo-differential operator \tilde{E}_2 of the symbol

$$\tilde{e}_2 = \overline{p}_m^{-1}(x, y, \xi, \eta) \# (1 + s_{-\min\{\epsilon_m, 1-r\}}(x, y, \xi, \eta)) \# \chi(x, y) (1 - \kappa(\xi, \eta)).$$

By construction we obtain $\tilde{P}\tilde{E}_2 = \chi(x, y)(1 - \kappa(D_x, D_y)) + \tilde{R}_2$ with \tilde{R}_2 regularizing on the anisotropic Sobolev spaces $H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta))$. Let $\tilde{E} = \tilde{E}_1 + \tilde{E}_2$. Then $\tilde{P}\tilde{E} = \tilde{P}\tilde{E}_1 + \tilde{P}\tilde{E}_2 = \chi(x, y) + \tilde{R}$, \tilde{R} regularizing.

After conjugation we obtain $PE = e^{-\tau\psi(y, D_x)}\chi(x, y)e^{\tau\psi(y, D_x)} + R$. Taking a function $\chi_0(x, y) \in G_0^{(q_1, q_2)}(\Omega)$ such that $\chi_0(x, y) = 1$ for $(x, y) \in \text{supp}(\chi)$ and replacing E by E_χ , where $E_\chi u := E(\chi(x, y)u)$, we have

$$\begin{aligned} P(x, y, D_x, D_y)E_\chi u &= e^{-\tau\psi(y, D_x)}\chi_0(x, y)e^{\tau\psi(y, D_x)}\chi(x, y)u + \bar{R}u \\ &= \chi(x, y)u - e^{-\tau\psi(y, D_x)}\tilde{R}_3e^{\tau\psi(y, D_x)}u + \bar{R}u. \end{aligned}$$

The operator $\tilde{R}_3 = (1 - \chi_0(x, y))e^{\tau\psi(y, D_x)}\chi(x, y)e^{-\tau\psi(y, D_x)}$ is again regularizing on the anisotropic Sobolev spaces $H_{\frac{d}{m}}^s(\mathbb{R} \times (-\delta, \delta))$ in view of Proposition 3.1. \square

Remark 3.3. Theorem 3.3, Remark 3.2, and Theorem 3.1, combined with fixed point arguments as in Gramchev and Rodino [GR, Section 4], lead to the proofs of Theorem 1.1 and Theorem 1.2.

Proof of Theorem 3.3. We limit ourselves for simplicity to proving the estimate (3.20) for $\alpha + \beta + \gamma + \theta = 1$. The case $\alpha + \beta + \gamma + \theta > 1$ does not involve actual complications; cf. Wakabayashi ([W], Theorem 2.6) or Kajitani and Wakabayashi ([KW], Theorem 1.9) for the analytic frame. First we estimate the numerator of (3.20), and then we give some lemmas to estimate the denominator.

If $\alpha = 1$, we get

$$\begin{aligned} &|D_x p(z, \zeta)|\lambda(\zeta)^{-\delta} \\ &\leq \left| \sum_{(l, j) \in I} D_x h_{lj}(x, y) \eta^j \xi^l - D_x h_{d0}(x, y) \xi^d \right| \lambda(\zeta)^{-\delta} + |D_x \sigma(z, \zeta)|\lambda(\zeta)^{-\delta} \\ &\leq L_1 \left(\left(\sum_{(l, j) \in I} |\eta|^j \xi^l + \xi^d \right) \lambda(\zeta)^{-\delta} + \lambda(\zeta)^{\bar{m}-\delta} \right); \end{aligned}$$

and for $\beta = 1$,

$$|D_y p(z, \zeta)|\lambda(\zeta)^{-\delta \frac{m}{d}} \leq L_2 \left(\left(\sum_{(l, j) \in I} |\eta|^j \xi^l + \xi^d \right) \lambda(\zeta)^{-\delta \frac{m}{d}} + \lambda(\zeta)^{\bar{m}-\delta \frac{m}{d}} \right),$$

for suitable constants L_1, L_2 . If $\gamma = 1$,

$$|D_\xi p(z, \zeta)|\lambda(\zeta)^{\rho \frac{m}{d}} \leq L_3 \left(\left(\sum_{(l, j) \in I} |\eta|^j \xi^{l-1} + \xi^{d-1} \right) \lambda(\zeta)^{\rho \frac{m}{d}} + \lambda(\zeta)^{\bar{m}-\frac{m}{d}(1-\rho)} \right);$$

and for $\theta = 1$,

$$|D_\eta p(z, \zeta)|\lambda(\zeta)^\rho \leq L_4 \left(\left(\sum_{(l, j) \in I} |\eta|^{j-1} \xi^l + |\eta|^{m-1} \right) \lambda(\zeta)^\rho + \lambda(\zeta)^{\bar{m}-(1-\rho)} \right),$$

with suitable constants L_3, L_4 .

Therefore, we observe that $\bar{m} - (1 - \rho) \geq \bar{m} - \frac{m}{d}(1 - \rho)$ since $d < m$, and $\bar{m} - (1 - \rho) = \bar{m} - \delta > \bar{m} - \frac{m}{d}\delta$ since $\rho + \delta = 1$. To prove (3.20), it will then be

sufficient to show the boundedness, for $|\zeta| > B$, of the functions

$$\begin{aligned} Q_1(\zeta) &= \frac{\left(\sum_{(l,j) \in I} |\eta|^j \xi^l + \xi^d\right) \lambda(\zeta)^{-\delta}}{|p(z, \zeta)|}, \\ Q_2(\zeta) &= \frac{\left(|\eta|^{m-1} + \sum_{(l,j) \in I} |\eta|^{j-1} \xi^l\right) \lambda(\zeta)^\rho}{|p(z, \zeta)|}, \\ Q_3(\zeta) &= \frac{\left(|\xi|^{d-1} + \sum_{(l,j) \in I} |\eta|^j \xi^{l-1}\right) \lambda(\zeta)^{\rho \frac{m}{d}}}{|p(z, \zeta)|}, \\ Q_4(\zeta) &= \frac{\lambda(\zeta)^{\overline{m} - (1-\rho)}}{|p(z, \zeta)|}. \end{aligned}$$

First we introduce three regions:

$$\begin{aligned} (3.24) \quad R_1 &: c|\xi|^d \leq |\eta|^m \leq C|\xi|^d, \\ R_2 &: |\eta|^m \geq C|\xi|^d, \\ R_3 &: |\eta|^m \leq c|\xi|^d, \end{aligned}$$

for suitable constants c, C satisfying $c < \frac{1}{2} \min_{(x,y) \in \Omega} |\Re h_{d0}(x, y)|$, and $C > 2 \max_{(x,y) \in \Omega} |\Re h_{d0}(x, y)|$. We understand the neighborhood Λ to be the region R_1 .

The following inequalities then hold:

$$(3.25) \quad \lambda(\zeta)^{-\delta} \leq \begin{cases} C^{\frac{\delta}{d}} |\eta|^{-\delta}, & \zeta \in R_1, & (I) \\ |\eta|^{-\delta}, & \zeta \in R_2, & (II) \\ |\xi|^{-\delta \frac{d}{m}}, & \zeta \in R_3, & (III) \end{cases}$$

and

$$\lambda(\zeta)^\rho \leq \begin{cases} C_1 |\eta|^\rho, & \zeta \in R_1, \\ C_2 |\eta|^\rho, & \zeta \in R_2, \\ C_3 |\xi|^{\rho \frac{d}{m}}, & \zeta \in R_3; \end{cases}$$

note that (II) and (III) in (3.25) hold for all $\zeta \in \mathbb{R}^2$, but for our aim we may limit ourselves to consider them respectively in R_2 and in R_3 . By abuse of notation, in the following we shall also denote by R_1, R_2, R_3 the sets $\Omega \times R_1, \Omega \times R_2, \Omega \times R_3$; recall that $\Gamma = \Omega \times \Lambda$.

Lemma 3.1. *Let $p(z, \zeta)$ be the function (3.16), such that (i), (ii), (iii) in (3.18) hold. Then there are positive constants $K_1 < 1$ and B , such that*

$$(3.26) \quad |p(z, \zeta)| \geq K_1 |\Im h_{l^*j^*}(x, y)| |\eta|^{j^*} |\xi|^{l^*}, \quad (z, \zeta) \in R_1, \quad |\zeta| > B.$$

Proof. We have

$$\begin{aligned} (3.27) \quad |p(z, \zeta)|^2 &= \left(\eta^m - \Re h_{d0}(x, y) \xi^d + \sum_{(l,j) \in I} \Re h_{lj}(x, y) \eta^j \xi^l \right. \\ &\quad \left. + \Re \sigma(z, \zeta) \right)^2 + \left(\Im h_{l^*j^*}(x, y) \eta^{j^*} \xi^{l^*} \right. \\ &\quad \left. + \sum_{(l,j) \in I_-} \Im h_{lj}(x, y) \eta^j \xi^l + \sum_{(l,j) \in I_+} \Im h_{lj}(x, y) \eta^j \xi^l \right. \\ &\quad \left. - \Im h_{d0}(x, y) \xi^d + \Im \sigma(z, \zeta) \right)^2. \end{aligned}$$

By removing the terms arising from the real part of $p(z, \zeta)$, we can write

$$|p(z, \zeta)|^2 \geq \Im h_{l^*j^*}(x, y)^2 \eta^{2j^*} \xi^{2l^*} + \sum_{i=1}^5 J_i(z, \zeta),$$

where

(3.28)

$$J_1 = \left(\sum_{(l,j) \in I_-} \Im h_{lj}(z) \eta^j \xi^l + \sum_{(l,j) \in I_+} \Im h_{lj}(z) \eta^j \xi^l - \Im h_{d0}(z) \xi^d + \Im \sigma(z, \zeta) \right)^2,$$

$$(3.29) \quad J_2(z, \zeta) = 2\Im h_{l^*j^*}(x, y) \sum_{(l,j) \in I_-} \Im h_{lj}(x, y) \eta^{j^*+j} \xi^{l^*+l},$$

$$(3.30) \quad J_3(z, \zeta) = 2\Im h_{l^*j^*}(x, y) \sum_{(l,j) \in I_+} \Im h_{lj}(x, y) \eta^{j^*+j} \xi^{l^*+l},$$

$$(3.31) \quad J_4(z, \zeta) = -2\Im h_{l^*j^*}(x, y) \Im h_{d0}(x, y) \eta^{j^*} \xi^{l^*+d},$$

$$(3.32) \quad J_5(z, \zeta) = 2\Im \sigma(z, \zeta) \Im h_{l^*j^*}(x, y).$$

The function (3.28) is nonnegative; (3.30) and (3.31) are also nonnegative by hypotheses (ii), (iii). Let us fix attention on $J_2(z, \zeta)$, defined by (3.29). We have, for all $\epsilon > 0$,

$$(\Im h_{l^*j^*}(x, y))^2 \eta^{2j^*} \xi^{2l^*} + J_2(z, \zeta) \geq (1 - \epsilon) (\Im h_{l^*j^*}(x, y))^2 \eta^{2j^*} \xi^{2l^*},$$

in R_1 , $|\zeta| > B$. In fact, by (3.24) in R_1 and hypothesis (i) in (3.18), for all $\epsilon > 0$ we get, for B sufficiently large,

$$\begin{aligned} \frac{|J_2(z, \zeta)|}{(\Im h_{l^*j^*}(x, y))^2 \eta^{2j^*} \xi^{2l^*}} &\leq \text{const} \sum_{(l,j) \in I_-} \frac{|\eta|^{j^*+j} |\xi|^{l^*+l}}{\eta^{2j^*} \xi^{2l^*}} \\ &\leq \text{const} \sum_{(l,j) \in I_-} \frac{|\eta|^{j^*+j+(l^*+l)\frac{m}{d}}}{\eta^{2j^*+2l^*\frac{m}{d}}} < \epsilon, \quad |\zeta| > B. \end{aligned}$$

We remark that $k^* = dj^* + ml^* > k^- = dj + ml$.

Concerning (3.32), since $\overline{m} < \frac{k^*}{d}$, we have

$$(\Im h_{l^*j^*}(x, y))^2 \eta^{2j^*} \xi^{2l^*} + J_5(z, \zeta) \geq (1 - \epsilon) (\Im h_{l^*j^*}(x, y))^2 \eta^{2j^*} \xi^{2l^*}.$$

Then

$$|p(z, \zeta)| \geq K_1 |\Im h_{l^*j^*}(x, y)| |\eta|^{j^*} |\xi|^{l^*}, \quad (z, \zeta) \in R_1, \quad |\zeta| > B,$$

for a suitable positive constant K_1 . □

Lemma 3.2. *Let $p(z, \zeta)$ be the function (3.16). Then there are positive constants $K_2 < 1$ and B , such that*

$$(3.33) \quad |p(z, \zeta)| \geq K_2 |\eta|^m, \quad (z, \zeta) \in R_2, \quad |\zeta| > B.$$

Proof. We write $|p(z, \zeta)|^2$ as in (3.27); by removing the terms arising from the imaginary part of $p(z, \zeta)$, we get

$$(3.34) \quad |p(z, \zeta)|^2 \geq (\eta^m - \Re h_{d0}(x, y) \xi^d)^2 + W_1(z, \zeta) + W_2(z, \zeta) + W_3(z, \zeta),$$

where

$$(3.35) \quad W_1(z, \zeta) = \left(\sum_{(l,j) \in I} \Re h_{lj}(x, y) \eta^j \xi^l + \Re \sigma(z, \zeta) \right)^2,$$

$$(3.36) \quad \begin{aligned} W_2(z, \zeta) = & 2 \sum_{(l,j) \in I} \Re h_{lj}(x, y) \eta^{j+m} \xi^l \\ & - 2 \Re h_{d0}(x, y) \sum_{(l,j) \in I} \Re h_{lj}(x, y) \eta^j \xi^{l+d}. \end{aligned}$$

$$(3.37) \quad W_3(z, \zeta) = 2\eta^m \Re \sigma(z, \zeta) - 2 \Re h_{d0}(x, y) \xi^d \sigma(z, \zeta).$$

Observe first that, for $\lambda > 0$ sufficiently small,

$$(\eta^m - \Re h_{d0}(x, y) \xi^d)^2 > \lambda \eta^{2m}.$$

In fact,

$$(\eta^m - \Re h_{d0}(x, y) \xi^d)^2 \geq \eta^{2m} - 2 \Re h_{d0}(x, y) \eta^m \xi^d,$$

and using (3.24) in R_2 , we have

$$\eta^{2m} - 2 \Re h_{d0}(x, y) \eta^m \xi^d \geq \left(1 - \frac{2}{C} \Re h_{d0}(x, y)\right) \eta^{2m} > \lambda \eta^{2m},$$

since $C > 2 \max_{(x,y) \in \Omega} |\Re h_{d0}(x, y)|$.

Equation (3.35) is nonnegative. We denote (3.36) by $\Upsilon_1(z, \zeta) - \Upsilon_2(z, \zeta)$ and (3.37) by $\Upsilon_3(z, \zeta) - \Upsilon_4(z, \zeta)$. Then

$$|p(z, \zeta)|^2 \geq \lambda \eta^{2m} + \Upsilon_1(z, \zeta) - \Upsilon_2(z, \zeta) + \Upsilon_3(z, \zeta) - \Upsilon_4(z, \zeta).$$

Arguing on $\Upsilon_1 - \Upsilon_2, \Upsilon_3 - \Upsilon_4$ in the same way as we did in Lemma 3.1, we can show that for all $\epsilon > 0$,

$$\lambda \eta^{2m} + \Upsilon_1(z, \zeta) - \Upsilon_2(z, \zeta) \geq (\lambda - \epsilon) \eta^{2m}, \quad (z, \zeta) \in R_2, |\zeta| > B,$$

and

$$\lambda \eta^{2m} + \Upsilon_3(z, \zeta) - \Upsilon_4(z, \zeta) \geq (\lambda - \epsilon) \eta^{2m}, \quad (z, \zeta) \in R_2, |\zeta| > B.$$

Thus

$$|p(z, \zeta)| \geq K_2 |\eta|^m, \quad (z, \zeta) \in R_2, |\zeta| > B,$$

where $K_2 = (\lambda - \epsilon)^{\frac{1}{2}}$. □

Lemma 3.3. *Let $p(z, \zeta)$ be the function (3.16), such that (iv) in (3.18) holds. Then there are positive constants $K_3 < 1$ and B such that*

$$(3.38) \quad |p(z, \zeta)| \geq K_3 |\xi|^d, \quad (z, \zeta) \in R_3, |\zeta| > B.$$

Proof. Again we apply (3.34), (3.35), (3.36), (3.37) to $|p(z, \zeta)|^2$. Observe that in R_3 , arguing as above, since $c < \frac{1}{2} \min_{(x,y) \in \Omega} |\Re h_{d0}(x, y)|$, we obtain, for a suitable constant $\mu > 0$,

$$(\eta^m - \Re h_{d0}(x, y)\xi^d)^2 > \mu \xi^{2d}.$$

About the terms in (3.35), (3.36) and (3.37), the remarks we made in Lemma 3.2 hold on replacing $\lambda \eta^{2m}$ with $\mu \xi^{2d}$. Then we have

$$|p(z, \zeta)| \geq K_3 |\xi|^d, \quad (z, \zeta) \in R_3, \quad |\zeta| > B,$$

where $K_3 = (\mu - \epsilon)^{\frac{1}{2}}$. □

We first consider $Q_1(\zeta)$ separately in the regions R_1, R_2, R_3 , to prove boundedness. In R_1 by (3.25), (3.26) we get easily, writing as before $k = dj + ml$,

$$Q_1(\zeta) \leq \text{const} \left(\sum_k \frac{1}{|\eta|^{m - \frac{k}{d}}} + 1 \right), \quad |\zeta| > B,$$

where $m - \frac{k}{d} > 0$ by definition of I and I_k . In the region R_2 we have $|p(z, \zeta)| \geq |\eta|^m > |\eta|^{\frac{k^*}{d}}$. In R_3 , by using (3.25) and (3.38) for a constant $\epsilon > 0$ which we may take as small as we want by fixing B sufficiently large, we have

$$Q_1(\zeta) \leq \text{const} \left(\sum_k \frac{1}{|\xi|^{2d - \frac{k}{m} - \frac{k^*}{m}}} + \frac{1}{|\xi|^{\delta \frac{d}{m}}} \right) < \epsilon, \quad |\zeta| > B.$$

We have therefore proved that $Q_1(\zeta)$ is bounded. Arguing in the same way on $Q_2(\zeta), Q_3(\zeta)$ and $Q_4(\zeta)$, we prove their boundedness in \mathbb{R}^2 .

Remark 3.4. By formulas (3.26), (3.33), (3.38), we obtain that $|p(z, \zeta)| \geq a |\zeta|^{\frac{k^*}{m}}$, $a > 0$, $|\zeta| > B$, since we are considering the case when $dj + ml < dm$. If we refer to the anisotropic weight function $\lambda(\zeta) = |\xi|^{\frac{d}{m}} + |\eta| \sim (|\xi|^d + |\eta|^m)^{\frac{1}{m}}$, we find that

$$(3.39) \quad |p(z, \zeta)| \geq b \lambda(\zeta)^{\frac{k^*}{d}}, \quad |\zeta| > B,$$

for a suitable positive constant b .

Now Lemma 3.1, Lemma 3.2, Lemma 3.3 and the estimate (3.39) complete the proof. □

Remark 3.5. It is possible to propose a geometric invariant generalization of Theorems 1.1 and 1.2, to pseudo-differential operators with involutive characteristics of multiplicity $m \geq 4$, in more than two space variables, by arguing microlocally, using classical Fourier integral operators and $S_{\rho, \delta}^m$ arguments. This will be the subject of another paper.

ACKNOWLEDGEMENT

The authors thank the unknown referee for critical remarks which led to improvements in the paper.

REFERENCES

- [BT] A. Bove and D. Tartakoff, *Propagation of Gevrey regularity for a class of hypoelliptic equations*, Trans. Amer. Math. Soc., **348** (1996), 2533-2575. MR **96i**:35017
- [BC] J. M. Bony and J. Y. Chemin, *Espaces fonctionnels associés au calcul de Weyl-Hörmander*, Bull. Soc. Math. France, **122** (1994), 77-118. MR **95a**:35152
- [CZ] M. Cicognani and L. Zanghirati, *On a class of unsolvable operators*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), **20** (1993), 357-369. MR **95d**:35003
- [C] A. Corli, *On local solvability of linear partial differential operators with multiple characteristics*, J. Differential Equations, **81** (1989), 275-293. MR **91e**:35007
- [C2] A. Corli, *On local solvability in Gevrey classes of linear partial differential operators with multiple characteristics*, Comm. Partial Differential Equations, **14** (1989), 1-25. MR **90a**:35001
- [DR2] G. De Donno and L. Rodino, *Gevrey hypoellipticity for partial differential equations with characteristics of higher multiplicity*, to appear in Rend. Sem. Mat. Univ. Politecnico Torino, (2000).
- [DR3] G. De Donno and L. Rodino, *Gevrey hypoellipticity for equations with involutive characteristics of higher multiplicity*, C. R. Acad. Bulgare Sci., **53** No. 7, (2000), 25-30. MR **2001g**:35045
- [D] B. Dehman, *Résolubilité local pour des équations semi-linéaires complexes*, Canad. J. Math., **42** (1990), 126-140. MR **91h**:35008
- [ES] Y. V. Egorov and D. W. Schulze, *Pseudo-differential operators, singularities, applications*, Operator Theory: Advances and Applications, Vol. 93, Birkhäuser-Verlag, Basel-Boston-Berlin, 1997. MR **98e**:35181
- [G] G. Garello, *Inhomogeneous paramultiplication and microlocal singularities for semilinear equations*, Boll. Un. Mat. Ital. (7) **10-B** (1996), 885-902. MR **97k**:35007
- [G1] G. Garello, *Local solvability for semilinear equations with multiple characteristics*, Ann. Univ. Ferrara Sez. VII, Sci. Mat. **41** (1996), suppl., 199-209. MR **98i**:35003
- [GG] T. Gramchev, *On the critical index of Gevrey solvability for some linear partial differential equations*, Workshop on Partial Differential Equations (Ferrara 1999), Ann. Univ. Ferrara Sez. VII (N.S.), **45** (1999), suppl. (2000), 139-153. MR **2002f**:35006
- [GP] T. Gramchev and P. Popivanov, *Local solvability of semilinear partial differential equations*, Ann. Univ. Ferrara Sez. VII - Sc. Mat. **35** (1989), 147-154. MR **91m**:35006
- [GP1] T. Gramchev and P. Popivanov, *Partial differential equations: Approximate solutions in scales of functional spaces*, Mathematical Research, **108**, Wiley-VCH Verlag, Berlin, 2000. MR **2001g**:35002
- [GPY] T. Gramchev, P. Popivanov, and M. Yoshino, *Critical Gevrey index for hypoellipticity of parabolic operators and Newton polygons*, Ann. Mat. Pura Appl., **170** (1996), 103-131. MR **98c**:35029
- [GR] T. Gramchev and L. Rodino, *Gevrey solvability for semilinear partial differential equations with multiple characteristics*, Boll. Un. Mat. Ital. B (8), **2** (1999), 65-120. MR **2001j**:35006
- [H] L. Hörmander, *The analysis of linear partial differential operators*, I, II, III, IV, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, 1983-1985, Berlin. MR **85g**:35002a; MR **85g**:35002b; MR **87d**:35002a; MR **87d**:35002b
- [HS] J. Hounie and P. Santiago, *On the local solvability of semilinear equations*, Comm. in Partial Differential Equations, **20** (1995), 1777-1789. MR **96h**:35005
- [HP] C. Hunt and A. Piriou, *Opérateurs pseudo-différentiels anisotropes d'ordre variable*, C. R. Acad. Sci. Paris, **268** (1969), 28-31. MR **40**:1819
- [HP1] C. Hunt and A. Piriou, *Majorations L^2 et inégalité sous-elliptique pour les opérateurs pseudo-différentiels anisotropes d'ordre variable*, C. R. Acad. Sci. Paris, **268** (1969), 214-217. MR **40**:1820

- [KS] K. Kajitani and S. Spagnolo, in progress, communicated to the meeting "Perturbative methods for nonlinear partial differential equations", Cagliari 2000.
- [KW] K. Kajitani and S. Wakabayashi, *Hypoelliptic operators in Gevrey classes*, in "Recent developments in hyperbolic equations" L. Cattabriga, F. Colombini, M.K.V. Murthy, S. Spagnolo, editors, Longman 1988, London, 115-134. MR **90e**:35041
- [LA] R. Lascar, *Distributions intégrale de Fourier et classes de Denjoy-Carleman. Applications*, C. R. Acad. Sci. Paris, Sér. A **284**, (1977), 485-488. MR **55**:906
- [L] H. Lewy, *An example of a smooth linear partial differential equation without solution*, Ann. of Math. (2), **66** (1957), 155-158. MR **19**:551d
- [LR1] O. Liess and L. Rodino, *Inhomogeneous Gevrey classes and related pseudo-differential operators*, Boll. Un. Mat. Ital., Sez. IV, **3-C** (1984), 233-323. MR **85k**:35239
- [LR2] O. Liess and L. Rodino, *Linear partial differential equations with multiple involutive characteristics*, in "Microlocal analysis and spectral theory", L. Rodino, editor, Kluwer, 1997, Dordrecht, 1-38. MR **98e**:35034
- [LO] M. Lorenz, *Anisotropic operators with characteristics of constant multiplicity*, Math. Nachr., **124** (1985), 199-216. MR **87f**:35249
- [MA] P. Marcolongo, *Solvability and nonsolvability for partial differential equations in Gevrey spaces*, Ph.D. Dissertation, Mathematics, University of Torino, 2000.
- [MO] P. Marcolongo and A. Oliaro, *Local solvability for semilinear anisotropic partial differential equations*, Ann. Mat. Pura e Appl. (4) **170** (2001), 229-262. MR **2002h**:35004
- [MR] M. Mascarello and L. Rodino, *Partial differential equations with multiple characteristics*, Wiley-VCH, 1997, Berlin. MR **99a**:35009
- [M] A. Menikoff, *On hypoelliptic operators with double characteristics*, Ann. Scuola Norm. Sup. Pisa, Cl. Sci. (4) **4** (1977), 689-724. MR **57**:13156
- [P1] P. R. Popivanov, *On the local solvability of a certain class of pseudo-differential equations with double characteristics*, Trudy Sem. Petrovsk., **1** (1975), 237-278; Amer. Math. Soc. Transl., **118** (1982), 51-90. MR **55**:841
- [P2] P. R. Popivanov, *Local solvability of some classes of linear differential operators with multiple characteristics*, Ann. Univ. Ferrara, Seg. VII, Sci. Mat. **45** suppl. (1999), 263-274. MR **2001j**:35038
- [P3] P. R. Popivanov, *Microlocal properties of a class of pseudodifferential operators with double involutive characteristics*, Partial Differential Equations, Banach Center Publications, Volume 19, PWN-Polish Scientific Publishers, Warsaw, 1987, pp. 213-224. MR **91i**:35221
- [PP] P. R. Popivanov and G. S. Popov, *Microlocal properties of a class of pseudo-differential operators with multiple characteristics*, Serdica, **6** (1980), 167-181. (Russian) MR **82d**:35038
- [R] G. B. Roberts, *Quasi-subelliptic estimates for operators with multiple characteristics*, Comm. Partial Differential Equations, **11** (1986), 231-320. MR **87e**:35024
- [RO] L. Rodino, *Linear partial differential operators in Gevrey spaces*, World Scientific Publishing Co., River Edge, NJ, 1993. MR **95c**:35001
- [RO2] L. Rodino, *Local solvability in Gevrey classes*, in: Hyperbolic Equations (Padua 1985), 167-185, Pitman Research Notes in Math. Ser., 158, Longman, Harlow, 1987. MR **89d**:35001
- [S] N. A. Šananin, *The local solvability of equations of quasi-principal type*, Mat. Sb. (N.S.) **97** (**139**), (1975), 503-516; English transl., Math. USSR Sb. **26** (1975), 458-470. MR **57**:13112
- [SE] F. Segala, *A class of locally solvable differential operators*, Boll. Un. Mat. Ital. B (6), **4** (1985), 241-251. MR **86i**:35003
- [SP] S. Spagnolo, *Local and semi-global solvability for systems of non-principal type*, Comm. Partial Differential Equations, **25**, no. 5-6, (2000), 1115-1141. MR **2002d**:35231
- [T] F. Trèves, *Introduction to pseudodifferential and Fourier integral operators*. I, II, The University Series in Mathematics, Plenum Press, 1980, New York and London. MR **82i**:35173; MR **82i**:58068
- [TU] V. N. Tulovsky, *Propagation of singularities of operators with characteristics of constant multiplicity*, Trudy Moskov. Mat. Obshch., **39** (1979), 113-134; English transl., Trans. Moscow Math. Soc., **1981**, no. 1 (39), 121-144. MR **82m**:35150

- [W] S. Wakabayashi, Singularities of solutions of the Cauchy problem for hyperbolic systems in Gevrey classes, Japan J. Math., **11** (1985), 157-201. MR **88h**:35067

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI TORINO, VIA CARLO ALBERTO 10, 10123 TORINO, ITALY

E-mail address: dedonno@dm.unito.it

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DI TORINO, VIA CARLO ALBERTO 10, 10123 TORINO, ITALY

E-mail address: oliaro@dm.unito.it

Editorial Information

To be published in the *Transactions*, a paper must be correct, new, nontrivial, and significant. Further, it must be well written and of interest to a substantial number of mathematicians. Piecemeal results, such as an inconclusive step toward an unproved major theorem or a minor variation on a known result, are in general not acceptable for publication.

Papers submitted to the *Transactions* should exceed 10 published journal pages in length. Shorter papers may be submitted to the *Proceedings of the American Mathematical Society*. Published pages are the same size as those generated in the style files provided for $\text{\AA MS-L}^{\text{\AA T}}\text{E}^{\text{\AA X}}$ or $\text{\AA MS-T}^{\text{\AA X}}$.

As of March 31, 2003, the backlog for this journal was approximately 2 issues. This estimate is the result of dividing the number of manuscripts for this journal in the Providence office that have not yet gone to the printer on the above date by the average number of articles per issue over the previous twelve months, reduced by the number of issues published in four months (the time necessary for editing and composing a typical issue). In an effort to make articles available as quickly as possible, articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

A Consent to Publish and Copyright Agreement is required before a paper will be published in this journal. After a paper is accepted for publication, the Providence office will send a Consent to Publish and Copyright Agreement to all authors of the paper. By submitting a paper to this journal, authors certify that the results have not been submitted to nor are they under consideration for publication by another journal, conference proceedings, or similar publication.

Information for Authors

Initial submission. Two copies of the paper should be sent directly to the appropriate Editor and the author should keep a copy. *If an editor is agreeable*, an electronic manuscript prepared in $\text{T}^{\text{E}}\text{X}$ or $\text{L}^{\text{A}}\text{T}^{\text{E}}\text{X}$ may be submitted by pointing to an appropriate URL on a preprint or e-print server.

The first page must consist of a *descriptive title*, followed by an *abstract* that summarizes the article in language suitable for workers in the general field (algebra, analysis, etc.). The *descriptive title* should be short, but informative; useless or vague phrases such as “some remarks about” or “concerning” should be avoided. The *abstract* should be at least one complete sentence, and at most 300 words. Included with the footnotes to the paper should be the 2000 *Mathematics Subject Classification* representing the primary and secondary subjects of the article. The classifications are accessible from www.ams.org/msc/. The list of classifications is also available in print starting with the 1999 annual index of *Mathematical Reviews*. The Mathematics Subject Classification footnote may be followed by a list of *key words and phrases* describing the subject matter of the article and taken from it. Journal abbreviations used in bibliographies are listed in the latest *Mathematical Reviews* annual index. The series abbreviations are also accessible from www.ams.org/publications/. To help in preparing and verifying references, the AMS offers MR Lookup, a Reference Tool for Linking, at www.ams.org/mrlookup/. When the manuscript is submitted, authors should supply the editor with electronic addresses if available. These will be printed after the postal address at the end of each article.

Electronically prepared manuscripts. The AMS encourages electronically prepared manuscripts, with a strong preference for $\text{\AA MS-L}^{\text{\AA T}}\text{E}^{\text{\AA X}}$. To this end, the

Society has prepared $\mathcal{AMS}\text{-}\text{\LaTeX}$ author packages for each AMS publication. Author packages include instructions for preparing electronic manuscripts, the *AMS Author Handbook*, samples, and a style file that generates the particular design specifications of that publication series. Articles properly prepared using the $\mathcal{AMS}\text{-}\text{\LaTeX}$ style file and the `\label` and `\ref` commands automatically enable extensive intra-document linking to the bibliography and other elements of the article for searching electronically on the Web. Because linking must often be added manually to electronically prepared manuscripts in other forms of \TeX , using $\mathcal{AMS}\text{-}\text{\LaTeX}$ also reduces the amount of technical intervention once the files are received by the AMS. This results in fewer errors in processing and saves the author proofreading time. $\mathcal{AMS}\text{-}\text{\LaTeX}$ papers also move more efficiently through the production stream, helping to minimize publishing costs.

$\mathcal{AMS}\text{-}\text{\LaTeX}$ is the highly preferred format of \TeX , but author packages are also available in $\mathcal{AMS}\text{-}\text{\TeX}$. Those authors who make use of these style files from the beginning of the writing process will further reduce their own efforts. Manuscripts prepared electronically in \LaTeX or plain \TeX are normally not acceptable due to the high amount of technical time required to insure that the file will run properly through the AMS in-house production system. \LaTeX users will find that $\mathcal{AMS}\text{-}\text{\LaTeX}$ is the same as \LaTeX with additional commands to simplify the typesetting of mathematics, and users of plain \TeX should have the foundation for learning $\mathcal{AMS}\text{-}\text{\LaTeX}$.

Authors may retrieve an author package from the AMS website starting from www.ams.org/tex/ or via FTP to [ftp.ams.org](ftp://ftp.ams.org) (login as `anonymous`, enter username as password, and type `cd pub/author-info`). The *AMS Author Handbook* and the *Instruction Manual* are available in PDF format following the author packages link from www.ams.org/tex/. The author package can also be obtained free of charge by sending email to pub@ams.org (Internet) or from the Publication Division, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When requesting an author package, please specify $\mathcal{AMS}\text{-}\text{\LaTeX}$ or $\mathcal{AMS}\text{-}\text{\TeX}$, Macintosh or IBM (3.5) format, and the publication in which your paper will appear. Please be sure to include your complete mailing address.

At the time of submission, authors should indicate if the paper has been prepared using $\mathcal{AMS}\text{-}\text{\LaTeX}$ or $\mathcal{AMS}\text{-}\text{\TeX}$ and provide the Editor with a paper manuscript that matches the electronic manuscript. The final version of the electronic manuscript should be sent to the Providence office immediately after the paper has been accepted for publication. The author should also send the final version of the paper manuscript to the Editor, who will forward a copy to the Providence office. Editors will require authors to send their electronically prepared manuscripts to the Providence office in a timely fashion. Electronically prepared manuscripts can be sent via email to pub-submit@ams.org (Internet) or on diskette to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When sending a manuscript electronically, please be sure to include a message indicating in which publication the paper has been accepted. No corrections will be accepted electronically. Authors must mark their changes on their proof copies and return them to the Providence office. Complete instructions on how to send files are included in the author package.

Electronic graphics. Comprehensive instructions on preparing graphics are available starting from www.ams.org/jourhtml/authors.html. A few of the major requirements are given here.

Submit files for graphics as EPS (Encapsulated PostScript) files. This includes graphics originated via a graphics application as well as scanned photographs or

other computer-generated images. If this is not possible, TIFF files are acceptable as long as they can be opened in Adobe Photoshop or Illustrator. No matter what method was used to produce the graphic, it is necessary to provide a paper copy to the AMS.

Authors using graphics packages for the creation of electronic art should also avoid the use of any lines thinner than 0.5 points in width. Many graphics packages allow the user to specify a “hairline” for a very thin line. Hairlines often look acceptable when proofed on a typical laser printer. However, when produced on a high-resolution laser imagesetter, hairlines become nearly invisible and will be lost entirely in the final printing process.

Screens should be set to values between 15% and 85%. Screens which fall outside of this range are too light or too dark to print correctly. Variations of screens within a graphic should be no less than 10%.

AMS policy on making changes to articles after posting. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue. To preserve the integrity of electronically published articles, once an article is individually posted to the AMS website but not yet in an issue, changes cannot be made in place in the paper. However, an “Added after posting” section may be added to the paper right before the References when there is a critical error in the content of the paper. The “Added after posting” section gives the author an opportunity to correct this type of critical error before the article is put into an issue for printing and before it is then reposted with the issue. The “Added after posting” section remains a permanent part of the paper. The AMS does not keep author-related information, such as affiliation, current address, and email address, up to date after a paper is initially posted.

Once the article is assigned to an issue, even if the issue has not yet been posted to the AMS website, corrections may be made to the paper by submitting a traditional errata article to the Editor. The errata article will appear in a future print issue and will link back and forth on the web to the original article online.

Secure manuscript tracking on the Web and via email. Authors can track their manuscripts through the AMS journal production process using the personal AMS ID and Article ID printed in the upper right-hand corner of the Consent to Publish form sent to each author who publishes in AMS journals. Access to the tracking system is available from www.ams.org/mstrack/ or via email sent to mstrack-query@ams.org. To access by email, on the subject line of the message simply enter the AMS ID and Article ID. To track more than one manuscript by email, choose one of the Article IDs and enter the AMS ID and the Article ID followed by the word *all* on the subject line. An explanation of each production step is provided on the web through links from the manuscript tracking screen. Questions can be sent to tran-query@ams.org.

T_EX files available. Beginning with the January 1992 issue of the *Bulletin* and the January 1996 issues of *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS*, T_EX files can be downloaded from the AMS website, starting from www.ams.org/journals/. Authors without Web access may request their files at the address given below after the article has been published. For *Bulletin* papers published in 1987 through 1991 and for *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS* papers published in 1987 through 1995, T_EX files are available upon request for authors without Web access by sending email to file-request@ams.org or by contacting the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. The request should include the title of the paper, the

name(s) of the author(s), the name of the publication in which the paper has or will appear, and the volume and issue numbers if known. The \TeX file will be sent to the author making the request after the article goes to the printer. If the requestor can receive Internet email, please include the email address to which the file should be sent. Otherwise please indicate a diskette format and postal address to which a disk should be mailed. **Note:** Because \TeX production at the AMS sometimes requires extra fonts and macros that are not yet publicly available, \TeX files cannot be guaranteed to run through the author's version of \TeX without errors. The AMS regrets that it cannot provide support to eliminate such errors in the author's \TeX environment.

Inquiries. Any inquiries concerning a paper that has been accepted for publication that cannot be answered via the manuscript tracking system mentioned above should be sent to tran-query@ams.org or directly to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

Editors

The traditional method of submitting a paper is to send two hard copies to the appropriate editor. Subjects, and the editors associated with them, are listed below.

In principle the Transactions welcomes electronic submissions, and some of the editors, those whose names appear below with an asterisk (*), have indicated that they prefer them. Editors reserve the right to request hard copies after papers have been submitted electronically. Authors are advised to make preliminary inquiries to editors as to whether they are likely to be able to handle submissions in a particular electronic form.

Algebra and algebraic geometry, KAREN E. SMITH, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109 USA; e-mail: kesmith@umich.edu

Algebraic geometry, DAN ABRAMOVICH, Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215 USA; e-mail: abramovic@bu.edu

Algebraic topology and cohomology of groups, STEWART PRIDDY, Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730 USA; e-mail: priddy@math.nwu.edu

* **Combinatorics**, SERGEY FOMIN, Department of Mathematics, East Hall, University of Michigan, Ann Arbor, MI 48109-1109 USA; e-mail: fomin@umich.edu

Complex analysis and geometry, D. H. PHONG, Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027-0029 USA; e-mail: phong@math.columbia.edu

* **Differential geometry and global analysis**, LISA C. JEFFREY, Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3; e-mail: jeffrey@math.toronto.edu

Dynamical systems and ergodic theory, ROBERT F. WILLIAMS, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: bob@math.utexas.edu

* **Geometric analysis**, TOBIAS COLDING, Courant Institute, New York University, 251 Mercer Street, New York, NY 10012 USA; e-mail: colding@cims.nyu.edu

Geometric topology, knot theory, and hyperbolic geometry, ABIGAIL THOMPSON, Department of Mathematics, University of California, Davis, CA 95616-5224 USA; e-mail: thompson@math.ucdavis.edu

Harmonic analysis, ALEXANDER NAGEL, Department of Mathematics, University of Wisconsin, 480 Lincoln Drive, Madison, WI 53706-1313 USA; e-mail: nagel@math.wisc.edu

Harmonic analysis, representation theory, and Lie theory, ROBERT J. STANTON, Department of Mathematics, Ohio State University, 231 West 18th Avenue, Columbus, OH 43210-1174 USA; e-mail: stanton@math.ohio-state.edu

* **Logic**, THEODORE SLAMAN, Department of Mathematics, University of California, Berkeley, CA 94720-3840 USA; e-mail: slaman@math.berkeley.edu

Number theory, HAROLD G. DIAMOND, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801-2917 USA; e-mail: diamond@math.uiuc.edu

* **Ordinary differential equations, partial differential equations, and applied mathematics**, PETER W. BATES, Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027 USA; e-mail: bates@math.msu.edu

* **Partial differential equations**, PATRICIA E. BAUMAN, Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395 USA; e-mail: bauman@math.purdue.edu

* **Probability and statistics**, KRZYSZTOF BURDZY, Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350 USA; e-mail: burdzy@math.washington.edu

* **Real analysis and partial differential equations**, DANIEL TATARU, Department of Mathematics, University of California, Berkeley, CA 94720 USA; e-mail: tataru@math.berkeley.edu

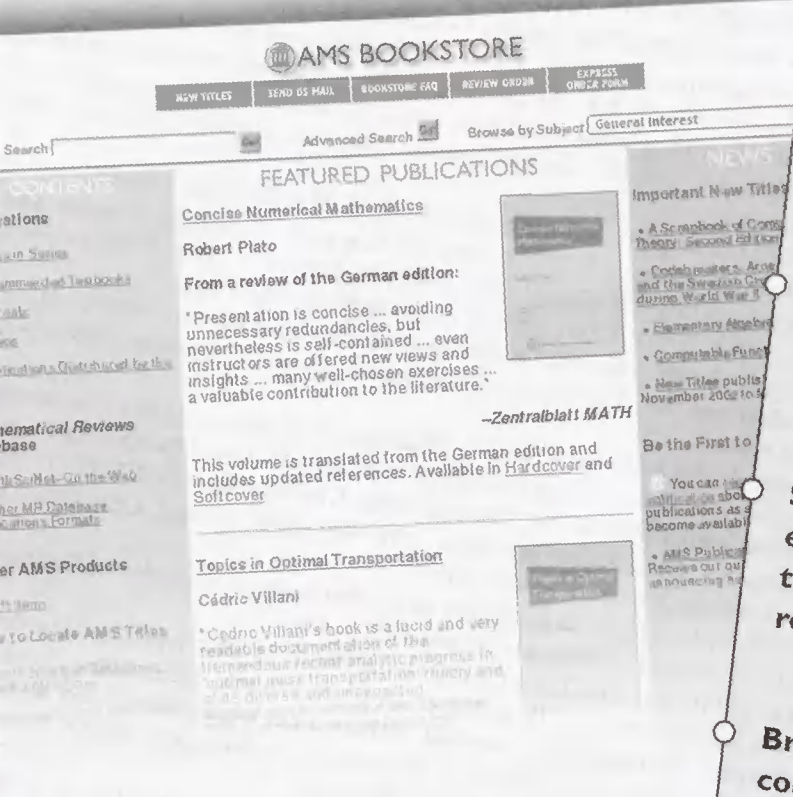
All other communications to the editors should be addressed to the Managing Editor, WILLIAM BECKNER, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: beckner@math.utexas.edu

MEMOIRS OF THE AMERICAN MATHEMATICAL SOCIETY

Memoirs is devoted to research in pure and applied mathematics of the same nature as *Transactions*. An issue consists of one or more separately bound research tracts for which the authors provide reproduction copy. Papers intended for *Memoirs* should normally be at least 80 pages in length. *Memoirs* has the same editorial committee as *Transactions*; so such papers should be addressed to one of the editors listed above.

Visit the AMS Bookstore

our key source for
AMS publications and purchases!



Check out the most up-to-date information on our books, journals, and electronic products and services.

Take advantage of web-exclusive sales.

Sign up to receive emails about new titles when they're released.

Browse the complete listing of over 3,000 books, journals, videos, gifts, and more.



amsbookstore.org



TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY

CONTENTS

Vol. 355, No. 8

Whole No. 819

August 2003

Robert Lauter and Sergiu Moroianu , Homology of pseudodifferential operators on manifolds with fibered cusps	3009
Yijun Hu and Tzong-Yow Lee , Moderate deviation principles for trajectories of sums of independent Banach space valued random variables	3047
Yanick Heurteaux , Weierstrass functions with random phases	3065
Stéphane Louboutin , Explicit lower bounds for residues at $s = 1$ of Dedekind zeta functions and relative class numbers of CM-fields	3079
Sophie Huczynska and Stephen D. Cohen , Primitive free cubics with specified norm and trace	3099
David Wright and Wenhua Zhao , D-log and formal flow for analytic isomorphisms of n -space	3117
Nobuo Hara and Ken-ichi Yoshida , A generalization of tight closure and multiplier ideals	3143
Jian Kong , Seshadri constants on Jacobian of curves	3175
Rajesh S. Kulkarni , On the Clifford algebra of a binary form	3181
Jaya N. Iyer , Projective normality of abelian varieties	3209
V. Braungardt and D. Kotschick , Clustering of critical points in Lefschetz fibrations and the symplectic Szpiro inequality	3217
Christian Wolf , On measures of maximal and full dimension for polynomial automorphisms of \mathbb{C}^2	3227
Mark Pollicott , Hausdorff dimension and asymptotic cycles	3241
Marius Dadarlat and Erik Guentner , Constructions preserving Hilbert space uniform embeddability of discrete groups	3253
Jaroslav Tišer , Vitali covering theorem in Hilbert space	3277
George B. Seligman , On idempotents in reduced enveloping algebras ...	3291
Hartmut Logemann, Richard Rebarber, and Stuart Townley , Stability of infinite-dimensional sampled-data systems	3301
Deguang Han , Approximations for Gabor and wavelet frames	3329
Tommaso Pacini , Mean curvature flow, orbits, moment maps	3343
Alexandru D. Ionescu , Singular integrals on symmetric spaces, II	3359
Sergey Antonyan , West's problem on equivariant hyperspaces and Banach-Mazur compacta	3379
Giuseppe de Donno and Alessandro Oliaro , Local solvability and hypoellipticity for semilinear anisotropic partial differential equations	3405



0002-9947(200308)355:8;1-I

VOLUME 355 NUMBER 9



SEPTEMBER

WHOLE NUMBER

TRANSACTION

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

Dan Abramovich

Peter W. Bates

Patricia E. Bauman

William Beckner, Managing Editor

Krzysztof Burdzy

Tobias Colding

Harold G. Diamond

Sergey Fomin

Lisa C. Jeffrey

Alexander Nagel

D. H. Phong

Stewart Priddy

Robert J. Stanton

Daniel Tataru

Robert F. Williams

PROVIDENCE, RHODE ISLAND USA

ISSN 0002-9947

Available electronically

www.ams.org

Transactions of the American Mathematical Society

This journal is devoted entirely to research in pure and applied mathematics.

Submission information. See **Information for Authors** at the end of this issue.

Publisher Item Identifier. The Publisher Item Identifier (PII) appears at the top of the first page of each article published in this journal. This alphanumeric string of characters uniquely identifies each article and can be used for future cataloging, searching, and electronic retrieval.

Postings to the AMS website. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

Subscription information. *Transactions of the American Mathematical Society* is published monthly. Beginning in January 1996 *Transactions* is accessible from www.ams.org/publications/. Subscription prices for Volume 355 (2003) are as follows: for paper delivery, \$1490 list, \$1192 institutional member, \$1341 corporate member; for electronic delivery, \$1341 list, \$1073 institutional member, \$1207 corporate member. Upon request, subscribers to paper delivery of this journal are also entitled to receive electronic delivery. If ordering the paper version, add \$39 for surface delivery outside the United States and India; \$50 to India. Expedited delivery to destinations in North America is \$48; elsewhere \$144. For paper delivery a late charge of 10% of the subscription price will be imposed upon orders received from nonmembers after January 1 of the subscription year.

Back number information. For back issues see www.ams.org/bookstore.

Subscriptions and orders should be addressed to the American Mathematical Society, P.O. Box 845904, Boston, MA 02284-5904 USA. *All orders must be accompanied by payment.* Other correspondence should be addressed to 201 Charles Street, Providence, RI 02904-2294 USA.

Copying and reprinting. Material in this journal may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. Requests can also be made by e-mail to reprint-permission@ams.org.

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

Transactions of the American Mathematical Society is published monthly by the American Mathematical Society at 201 Charles Street, Providence, RI 02904-2294 USA. Periodicals postage is paid at Providence, Rhode Island. Postmaster: Send address changes to *Transactions*, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

© 2003 by the American Mathematical Society. All rights reserved.

This journal is indexed in *Mathematical Reviews*, *Zentralblatt MATH*, *Science Citation Index*®, *Science Citation Index*™-Expanded, *ISI Alerting Services*™, *CompuMath Citation Index*®, and *Current Contents*®/Physical, Chemical & Earth Sciences.

Printed in the United States of America.

⊗ The paper used in this journal is acid-free and falls within the guidelines established to ensure permanence and durability.

Heinz H. Bauschke, Frank Deutsch, Hein Hundal, and Sung-Ho Park , Accelerating the convergence of the method of alternating projections	3433
Sunghan Bae, Ernst-Ulrich Gekeler, Pyung-Lyun Kang, and Linsheng Yin , Anderson's double complex and gamma monomials for rational function fields	3463
Lucia Caporaso , Remarks about uniform boundedness of rational points over function fields	3475
Thomas Keilen , Irreducibility of equisingular families of curves	3485
L. Brandolini, A. Iosevich, and G. Travaglini , Planar convex bodies, Fourier transform, lattice points, and irregularities of distribution	3513
Shangbin Cui and Avner Friedman , A free boundary problem for a singular system of differential equations: An application to a model of tumor growth	3537
José García-Cuerva, José Manuel Marco, and Javier Parcet , Sharp Fourier type and cotype with respect to compact semisimple Lie groups	3591
J. Rosický and W. Tholen , Left-determined model categories and universal homotopy theories	3611
Christian Henriksen , The combinatorial rigidity conjecture is false for cubic polynomials	3625
Leo T. Butler , Zero entropy, non-integrable geodesic flows and a non-commutative rotation vector	3641
Carlos Sancho de Salas , Complete homogeneous varieties: Structure and classification	3651
Neil O'Connell , A path-transformation for random walks and the Robinson-Schensted correspondence	3669
Takae Tsuji , On the Iwasawa λ -invariants of real abelian fields	3699
Xiaoxiang Jiao and Jiagui Peng , Pseudo-holomorphic curves in complex Grassmann manifolds	3715
Vassilis G. Papanicolaou , The periodic Euler-Bernoulli equation	3727
Francisco Jesús Castro-Jiménez and Nobuki Takayama , Singularities of the hypergeometric system associated with a monomial curve	3761
Jason P. Bell and Stanley N. Burris , Asymptotics for logical limit laws: When the growth of the components is in an RT class	3777
Michael E. Hoffman , Combinatorics of rooted trees and Hopf algebras	3795
Mahuya Datta , Connections with prescribed first Pontrjagin form	3813
Toru Ohmoto, Osamu Saeki, and Kazuhiro Sakuma , Self-intersection class for singularities and its application to fold maps	3825
A. Ülger , Erratum to "Arens regularity of the algebra $A \hat{\otimes} B$ "	3839
Nguyễn H. V. Hung , Erratum to "Spherical classes and the algebraic transfer"	3841

ACCELERATING THE CONVERGENCE OF THE METHOD OF ALTERNATING PROJECTIONS

HEINZ H. BAUSCHKE, FRANK DEUTSCH, HEIN HUNDAL, AND SUNG-HO PARK

ABSTRACT. The powerful von Neumann-Halperin method of alternating projections (MAP) is an algorithm for determining the best approximation to any given point in a Hilbert space from the intersection of a finite number of subspaces. It achieves this by reducing the problem to an iterative scheme which involves only computing best approximations from the *individual* subspaces which make up the intersection. The main practical drawback of this algorithm, at least for some applications, is that the method is slowly convergent. In this paper, we consider a general class of iterative methods which includes the MAP as a special case. For such methods, we study an “accelerated” version of this algorithm that was considered earlier by Gubin, Polyak, and Raik (1967) and by Gearhart and Koshy (1989). We show that the accelerated algorithm converges faster than the MAP in the case of two subspaces, but is, in general, *not faster* than the MAP for more than two subspaces! However, for a “symmetric” version of the MAP, the accelerated algorithm always converges faster for any number of subspaces. Our proof seems to require the use of the Spectral Theorem for selfadjoint mappings.

1. INTRODUCTION

Let X be a (real) Hilbert space, let M_1, M_2, \dots, M_k be closed (linear) subspaces of X with $M = \bigcap_1^k M_i$, and for any closed subspace N of X , let P_N denote the orthogonal projection onto N . The von Neumann-Halperin method of alternating projections, or MAP for short, is an iterative algorithm for determining the best approximation $P_M x$ to x from M . It does this by computing the iterates $x_0 := x$ and $x_n = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1}) x_{n-1} = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x$. That is, the iterates (x_n) are obtained by cyclically computing the best approximations onto the individual subspaces M_i ($i = 1, 2, \dots, k$). The method is thus most effective when it is “easy” to compute the best approximations from the individual subspaces M_i . The main theorem governing the MAP is the following.

Theorem (von Neumann [18] for $k = 2$, Halperin [15] for $k \geq 2$). *Let M_1, M_2, \dots, M_k be closed subspaces in the Hilbert space X and let $M := \bigcap_1^k M_i$. Then*

$$\lim_{n \rightarrow \infty} \|(P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x - P_M x\| = 0 \quad \text{for all } x \in X.$$

In case $k = 2$, this result was rediscovered in at least six other papers (see, e.g., the survey [5]).

Received by the editors July 30, 1999.

2000 *Mathematics Subject Classification.* Primary 41A65.

Key words and phrases. Alternating projections, cyclic projections, accelerating convergence, best approximation from an intersection of subspaces, Hilbert space.

Also, as was noted in [5], there are at least ten different areas of mathematics in which the MAP has proved useful. However, the main *practical* drawback of the MAP appears to be that it is often slowly convergent. Indeed, if $M_1 + M_2$ is not closed, then Franchetti and Light [11] and Bauschke, Borwein, and Lewis [2] have given examples showing that the convergence of $(P_{M_2}P_{M_1})^n x$ to $P_{M_1 \cap M_2} x$ can be arbitrarily slow!

Both Gubin, Polyak, and Raik [14] and Gearhart and Koshy [13] have considered a geometrically appealing method to *accelerate* the MAP, but in neither of these two papers was it proved that the acceleration scheme considered was actually faster than the MAP. In this paper, we will prove that this acceleration scheme is indeed faster than the MAP in the case of two subspaces (i.e., $k = 2$) (Theorem 3.23). But, perhaps surprisingly, we show that the acceleration scheme may actually be *slower* than the MAP when $k \geq 3$ (Example 3.24)! In contrast to this, we show that a “symmetric” version of the MAP (i.e., $x_0 = x$ and $x_n = (P_{M_1}P_{M_2} \cdots P_{M_k}P_{M_{k-1}} \cdots P_{M_1})^n x$ for $n = 1, 2, \dots$) has an accelerated version which is faster for any $k \geq 2$ (Corollary 3.21).

We should also mention that Dyer [10] and Hanke and Niethammer [16] have considered methods of accelerating the “Kaczmarz method” of solving linear equations. (Recall that Kaczmarz’s method may be regarded as the special case of the MAP in the case when X is finite-dimensional and each M_i is a hyperplane.)

2. THE METHOD OF ITERATED PROJECTIONS

To provide motivation for the acceleration results to be established later, in this section we give a fairly general convergence result which contains the von Neumann-Halperin result as a special case. In the next section, we will consider methods to accelerate this general algorithm.

Unless otherwise stated, the standing assumptions are as follows. Let X be a (real) Hilbert space, M_1, M_2, \dots, M_k be closed subspaces, $M := \bigcap_1^k M_i$, and let $P_i = P_{M_i}$ denote the orthogonal projection onto M_i ($i = 1, 2, \dots, k$).

Now let

$$T := P_k P_{k-1} \cdots P_1$$

denote the composition of the k projections P_i taken in increasing order. The well-known von Neumann-Halperin Theorem states that

$$\lim_{n \rightarrow \infty} \|T^n x - P_M x\| = 0$$

for each $x \in X$ (see, more generally, Theorem 2.5 below). Also, it can be shown that

$$\lim_n \|(T^* T)^n x - P_M x\| = 0$$

for each $x \in X$ (see Theorem 2.6 below). More generally, suppose T is any bounded linear mapping from X into itself such that

$$(2.0.1) \qquad \lim_n \|T^n x - P_{\text{Fix } T} x\| = 0 \quad \text{for each } x \in X,$$

where

$$\text{Fix } T := \{x \in X \mid Tx = x\}$$

is the *fixed point* set for T .

We will be interested in determining methods to *accelerate* the convergence of the sequence $(T^n x)$ to $P_{\text{Fix } T} x$. Before we consider such methods, it will provide

useful motivation to first give some general conditions on the mapping T that will guarantee that (2.0.1) holds.

The mapping T is called **nonexpansive** if $\|T\| \leq 1$. We first recall that the fixed point sets of T and T^* are the same if T is nonexpansive (see Riesz and Sz.-Nagy [19] or Riesz and Sz.-Nagy [20, p. 408]).

Lemma 2.1. *Let T be a nonexpansive linear operator on X . Then*

$$(2.1.1) \quad \text{Fix } T = \text{Fix } T^*.$$

In fact, $Tx = x$ if and only if $\langle Tx, x \rangle = \|x\|^2$ if and only if $\langle x, T^*x \rangle = \|x\|^2$ if and only if $T^*x = x$.

Our next observation is a characterization of those linear operators T on X that satisfy (2.0.1). We will use the following notation. If A is any linear operator on X , we denote the *range* and *null space* of A by $\mathcal{R}(A)$ and $\mathcal{N}(A)$ respectively. We will also use the well-known fact that $\mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$ (see [3, Remarks following Theorem 2.19 on pp. 35-36]).

Theorem 2.2. *Let T be a bounded linear operator on X , and let M be a closed linear subspace of X . Consider the following statements:*

- (1) $\lim_n \|T^n x - P_M x\| = 0$ for each $x \in X$;
- (2) $M = \text{Fix } T$ and $T^n x \rightarrow 0$ for each $x \in M^\perp$;
- (3) $M = \text{Fix } T$ and T is “asymptotically regular”, i.e., $T^n x - T^{n+1} x \rightarrow 0$ for each $x \in X$.

Then (1) \iff (2) \implies (3). If, in addition, T is nonexpansive, then all three statements are equivalent.

Proof. Suppose (1) holds. If $x \in M$, then $T^n x \rightarrow P_M x = x$. So by the continuity of T ,

$$Tx = T(\lim T^n x) = \lim T(T^n x) = \lim T^{n+1} x = P_M x = x$$

implies that $x \in \text{Fix } T$, i.e., $M \subset \text{Fix } T$.

Conversely, let $y \in \text{Fix } T$. Then $Ty = y$ and an easy induction shows that $y = T^n y$ for each n . Thus $y = T^n y \rightarrow P_M y$ which implies $y = P_M y \in M$. That is, $\text{Fix } T \subset M$. Hence $M = \text{Fix } T$.

If $x \in M^\perp$, then

$$T^n x = T^n(P_{M^\perp} x) \rightarrow P_M(P_{M^\perp} x) = 0.$$

This proves (2).

Now assume (2) holds and let $x \in X$. Then

$$T^n x = T^n(P_M x + P_{M^\perp} x) = T^n(P_M x) + T^n(P_{M^\perp} x) = P_M x + T^n(P_{M^\perp} x) \rightarrow P_M x.$$

Thus (1) holds, and this establishes the equivalence of (1) and (2).

Now suppose that (2) holds and $x \in X$. By the equivalence of (1) and (2), we have that $T^n x \rightarrow P_M x$ and so $T^n x - T^{n+1} x \rightarrow P_M x - P_M x = 0$. Thus T is asymptotically regular, and hence (3) holds.

This proves the first statement of the theorem. To complete the proof, suppose (3) holds and let T be nonexpansive. Then $\text{Fix } T^* = \text{Fix } T = M$ by Lemma 2.1. Then for any $x \in X$, we have that $T^n(x - Tx) = T^n x - T^{n+1} x \rightarrow 0$. Hence $T^n y \rightarrow 0$ for every $y \in \mathcal{R}(I - T)$ which implies, since $\|T^n\| \leq 1$ by nonexpansiveness, that $T^n y \rightarrow 0$ for every

$$y \in \overline{\mathcal{R}(I - T)} = \mathcal{N}(I - T^*)^\perp = (\text{Fix } T^*)^\perp = M^\perp.$$

Thus, for any $x \in X$,

$$T^n x = T^n(P_M x + P_{M^\perp} x) = T^n(P_M x) + T^n(P_{M^\perp} x) = P_M x + T^n(P_{M^\perp} x) \rightarrow P_M x,$$

and this proves that (1) holds. \square

Remark. Statement (3) does *not* imply statement (1) in general. To see this, let X denote the Euclidean plane and let $e_1 = (1, 0)$ and $e_2 = (0, 1)$ denote the canonical orthonormal basis vectors in X . Defining $T : X \rightarrow X$ by $Tx = [x(1) + x(2)]e_1$, it is easy to verify that $T^n x = Tx$ for every $n \in \mathbb{N}$ and every $x \in X$, so that T is asymptotically regular, $M := \text{Fix } T = \text{span } e_1$, $T^n(e_1 + e_2) = 2e_1$ for every n , but $P_M(e_1 + e_2) = e_1 \neq 2e_1 = T^n(e_1 + e_2)$ for every n . Thus, $T^n x \not\rightarrow P_M(x)$ when $x = e_1 + e_2$.

Corollary 2.3. *Let T be nonexpansive on X and $M = \text{Fix } T$. Then*

$$\lim_{n \rightarrow \infty} \|T^n x - P_M x\| = 0 \quad \text{for all } x \in X$$

if and only if T is asymptotically regular.

Lemma 2.4. *Let M_1, M_2, \dots, M_k be closed subspaces of the Hilbert space X , let $M := \bigcap_1^k M_i$ and let $T := P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$. Then T is nonexpansive and*

$$\text{Fix } T = \text{Fix } T^* = \text{Fix } (TT^*) = \text{Fix } (T^*T) = M.$$

Proof. For simplicity, let $P_i = P_{M_i}$. Since T is the product of nonexpansive operators, T is nonexpansive. If $x \in M$, then $x \in M_i$ for each i so that $P_i x = x$ for each i and hence $Tx = x$. That is, $M \subset \text{Fix } T$. Conversely, if $z \in \text{Fix } T$, then $Tz = z$. Thus, $P_k P_{k-1} \cdots P_1 z = z$. We have $P_i z = z$ if and only if $\|P_i z\| = \|z\|$ (using the fact that $\|z\|^2 = \|P_i z\|^2 + \|z - P_i z\|^2$). If $z \notin M$, let i be the smallest index such that $z \notin M_i$. Then $P_i z \neq z$; so $\|P_i z\| < \|z\|$ and $z = P_k \cdots P_i P_{i-1} \cdots P_1 z = P_k \cdots P_i z$ implies that $\|z\| = \|P_k \cdots P_i z\| \leq \|P_i z\| < \|z\|$, which is absurd. Thus, $z \in M$. This proves that $M = \text{Fix } T$. By Lemma 2.1, $M = \text{Fix } T^*$.

Since $T^* = P_1 P_2 \cdots P_k$, we see that $TT^* = P_k P_{k-1} \cdots P_1 P_2 \cdots P_k$ and $T^*T = P_1 P_2 \cdots P_k P_{k-1} \cdots P_1$, and the same proof as above shows that $\text{Fix } TT^* = M = \text{Fix } T^*T$. \square

A useful sufficient condition that guarantees that (2.0.1) holds is essentially contained in Halperin [15]. It also is explicit in Smarzewski [21] and can be stated as follows. (We include a brief proof since, as far as we know, the paper [21] has not been published.) Recall that $T : X \rightarrow X$ is called **nonnegative** if $\langle Tx, x \rangle \geq 0$ for all $x \in X$.

Theorem 2.5. *Let T_1, T_2, \dots, T_k be selfadjoint, nonnegative, and nonexpansive bounded linear operators on the Hilbert space X . Let $T := T_1 T_2 \cdots T_k$ and $M = \text{Fix } T$. Then $\text{Fix } T = \bigcap_1^k \text{Fix } T_i$ and*

$$(2.5.1) \quad \lim_n \|T^n x - P_M x\| = 0 \quad \text{for every } x \in X.$$

Proof. Since T is nonexpansive, Corollary 2.3 implies that it suffices to show that T is asymptotically regular. Toward this end, note that for each i , $I - T_i$ is nonnegative (and selfadjoint) since

$$\langle (I - T_i)x, x \rangle = \langle x - T_i x, x \rangle = \|x\|^2 - \langle T_i x, x \rangle \geq \|x\|^2 - \|T_i\| \|x\|^2 \geq 0.$$

It follows from a result of Riesz (see [4, Theorem 4.6.4, p. 163]) that $T_i(I - T_i)$ is also nonnegative. Hence,

$$\begin{aligned}\|x\|^2 &= \|x - T_i x + T_i x\|^2 = \|x - T_i x\|^2 + 2\langle x - T_i x, T_i x \rangle + \|T_i x\|^2 \\ &= \|x - T_i x\|^2 + 2\langle T_i(I - T_i)x, x \rangle + \|T_i x\|^2 \geq \|x - T_i x\|^2 + \|T_i x\|^2.\end{aligned}$$

Thus, for each $x \in X$,

$$(2.5.2) \quad \|x - T_i x\|^2 \leq \|x\|^2 - \|T_i x\|^2 \quad \text{for each } i.$$

By repeated application of (2.5.2), we deduce that

$$\begin{aligned}\|x\|^2 - \|Tx\|^2 &= \|x\|^2 - \|T_k x\|^2 + \|T_k x\|^2 - \|T_{k-1} T_k x\|^2 + \cdots + \|T_2 \cdots T_k x\|^2 - \|Tx\|^2 \\ &\geq \|x - T_k x\|^2 + \|T_k x - T_{k-1} T_k x\|^2 + \cdots + \|T_2 \cdots T_k x - Tx\|^2 \\ &= k \left[\frac{1}{k} \|x - T_k x\|^2 + \frac{1}{k} \|T_k x - T_{k-1} T_k x\|^2 + \cdots + \frac{1}{k} \|T_2 \cdots T_k x - Tx\|^2 \right] \\ &\geq k \left\| \frac{1}{k} (x - T_k x + T_k x - T_{k-1} T_k x + \cdots + T_2 \cdots T_k x - Tx) \right\|^2 \\ &\quad (\text{by convexity of } \|\cdot\|^2) \\ &= \frac{1}{k} \|x - Tx\|^2.\end{aligned}$$

That is,

$$(2.5.3) \quad \|x - Tx\|^2 \leq k(\|x\|^2 - \|Tx\|^2) \quad \text{for every } x \in X.$$

Since T is nonexpansive, we see that the sequence $(\|T^n x\|)_{n=1}^\infty$ is nonincreasing for every $x \in X$ and so it must converge: $\|T^n x\| \rightarrow \rho \geq 0$. Now apply (2.5.3) with x replaced by $T^n x$ to obtain that

$$\|T^n x - T^{n+1} x\|^2 \leq k(\|T^n x\|^2 - \|T^{n+1} x\|^2) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This proves that T is asymptotically regular. \square

Lemma 2.4 and Theorem 2.5 immediately imply the following two results. The first is the “von Neumann-Halperin theorem” stated in the Introduction, while the second shows that a symmetric version of the MAP also converges.

Theorem 2.6. *Let M_1, M_2, \dots, M_k be closed subspaces of the Hilbert space X , and let $M = \bigcap_{i=1}^k M_i$. Then, for each $x \in X$,*

$$(2.6.1) \quad \lim_n \|(P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x - P_M x\| = 0.$$

Theorem 2.7. *Let M_1, M_2, \dots, M_k be closed subspaces of the Hilbert space X , and let $M = \bigcap_{i=1}^k M_i$. Then, for each $x \in X$,*

$$(2.7.1) \quad \lim_n \|(P_{M_1} P_{M_2} \cdots P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n x - P_M x\| = 0.$$

Using Theorems 2.6 and 2.7, we see that two important examples of operators T which satisfy (2.0.1) are $T = Q$ and $T = Q^*Q$, where $Q := P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$.

3. ACCELERATION METHODS

Throughout this section, unless explicitly stated otherwise, we assume that T is a nonexpansive linear operator on X and $M := \text{Fix } T$. Hence, $\text{Fix } T^* = M$ also (by Lemma 2.1). Moreover, M_i will always denote a closed linear subspace of X and $P_i = P_{M_i}$.

In this section, we develop our main results concerned with accelerating the method given by (2.0.1). That is, if T is an operator such that (2.0.1) holds (or equivalently, that T is asymptotically regular), how can we modify the iterates suggested by this algorithm so as to converge faster to $P_M x$?

Definition 3.1. The **accelerated mapping** A_T of T is defined on X by

$$(3.1.1) \quad A_T(x) := t_x T x + (1 - t_x)x,$$

where

$$(3.1.2) \quad t_x = t_{x,T} := \begin{cases} \frac{\langle x, x - T x \rangle}{\|x - T x\|^2} & \text{if } T x \neq x, \\ 1 & \text{if } T x = x. \end{cases}$$

We will consider two classes of iterative algorithms to compute $P_M(x)$ for any given $x \in X$. They are described as follows. The standard or “unaccelerated” algorithm: $x_0 = x$ and

$$(3.1.3) \quad x_n = T(x_{n-1}) = T^n(x) \quad (n = 1, 2, \dots),$$

and its “accelerated” counterpart: $x_0 = x$, $x_1 = T(x_0)$, and

$$(3.1.4) \quad x_n = A_T(x_{n-1}) = A_T^{n-1}(T x) \quad (n = 1, 2, \dots).$$

In particular, we will give a detailed analysis of these algorithms when $T = P_k P_{k-1} \cdots P_1$ and when $T = (P_k P_{k-1} \cdots P_1)^*(P_k P_{k-1} \cdots P_1)$. This acceleration scheme was suggested by Gubin et al [14] and Gearhart and Koshy [13] in the particular case when T is a product of projections. The motivation for using the mapping A_T is that $A_T(x)$ is that point on the line through the points x and $T x$ which is closest to $P_M x$ (see Theorem 3.7 below).

A remark is in order as to why, in the accelerated algorithm, we first apply T to x_0 rather than first applying A_T . That is, why didn't we define the accelerated algorithm by $x_n = A_T^n(x_0)$ for $n \geq 0$ rather than $x_{n+1} = A_T^n(T x_0)$ for $n \geq 0$? The simple answer is that, besides being the one suggested in [14] and [13], the one we defined performs better. Indeed, it is not hard to see that if T is the product of two orthogonal projections onto two 1-dimensional (nonorthogonal) subspaces in the Euclidean plane, then *the accelerated algorithm converges in two steps*, that is, $A_T(T x) = P_M x$ for any starting point x . However, for any choice of x which is not in the range of T , *none* of the terms of the sequence $(A_T^n(x))$ is equal to $P_M x$. That is, the sequence $x_n = A_T^n(x)$ does not converge to $P_M x$ in a finite number of steps.

Definition 3.2. The classical **von Neumann-Halperin method of alternating projections**, or **MAP** for short, corresponds to (3.1.3) in the case when $T = P_k P_{k-1} \cdots P_1$.

The **accelerated method of alternating projections**, or the **accelerated MAP** for short, is the algorithm (3.1.4) in the case when $T = P_k P_{k-1} \cdots P_1$.

The **symmetric method of alternating projections**, or **symmetric MAP** for short, is just (3.1.3) in the case when $T = (P_k P_{k-1} \cdots P_1)^*(P_k P_{k-1} \cdots P_1)$.

The **accelerated symmetric method of alternating projections**, or **accelerated symmetric MAP** for short, is the algorithm (3.1.4) in the case when $T = (P_k P_{k-1} \cdots P_1)^* (P_k P_{k-1} \cdots P_1)$.

Lemma 3.3. *Let $x \in X$. Then*

- (1) $tA_T(x) + (1-t)x - x \in M^\perp \cap (A_T(x))^\perp$ for every $t \in \mathbb{R}$.
- (2) $Tx - x \in M^\perp \cap (A_T(x))^\perp$.
- (3) $A_T(x) - x \in M^\perp \cap (A_T(x))^\perp$.
- (4) $T(M^\perp) \subset M^\perp$ and $A_T(M^\perp) \subset M^\perp$.
- (5) $A_T(x) - Tx \in M^\perp \cap (A_T(x))^\perp$.
- (6) $Tx - P_M x \in M^\perp$.

Proof. (1) Since $tA_T(x) + (1-t)x - x = tt_x(Tx - x)$, it suffices to verify (2).

(2) If $Tx = x$, then (2) is trivial. Thus we may assume that $Tx \neq x$. Let $y \in M$. Then since $Ty = y = T^*y$, we have

$$\langle Tx - x, y \rangle = \langle x, T^*y \rangle - \langle x, y \rangle = \langle x, y \rangle - \langle x, y \rangle = 0.$$

Thus $Tx - x \in M^\perp$. Also,

$$\begin{aligned} \langle Tx - x, A_T(x) \rangle &= \langle Tx - x, t_x Tx + (1 - t_x)x \rangle = t_x \langle Tx - x, Tx - x \rangle + \langle Tx - x, x \rangle \\ &= -\frac{\langle x, Tx - x \rangle}{\|x - Tx\|^2} \|Tx - x\|^2 + \langle Tx - x, x \rangle = 0; \end{aligned}$$

so $Tx - x \in (A_T(x))^\perp$.

- (3) Take $t = 1$ in (1).
- (4) This follows from (2) and (3).
- (5) Since $A_T(x) - Tx = (t_x - 1)(Tx - x)$, the result follows from (2).
- (6) Using part (2), we get

$$Tx - P_M x = (Tx - x) + (x - P_M x) \in M^\perp + M^\perp = M^\perp.$$

□

Lemma 3.4. *For every $x \in X$ and $n \in \mathbb{N} := \{1, 2, \dots\}$,*

- (1) $P_M(A_T(x)) = P_M x$;
- (2) $P_M(Tx) = P_M x$;
- (3) $P_M(A_T^{n-1}(Tx)) = P_M x$;
- (4) $P_M(T^n x) = P_M x$.

Proof. We use the well-known fact that $P_M(M^\perp) = \{0\}$. Since $Tx - x \in M^\perp$ and $A_T(x) - x \in M^\perp$ by Lemma 3.3, it follows that $0 = P_M(Tx - x) = P_M(Tx) - P_M x$ and $0 = P_M(A_T(x) - x) = P_M(A_T(x)) - P_M x$. Hence (1) and (2) follow.

(3) and (4) follow by a repeated application of (1) and (2). □

Lemma 3.5. *For each $x \in X$ and $y \in M$,*

$$(3.5.1) \quad \|A_T(x) - y\|^2 = \|x - y\|^2 - \|x - A_T(x)\|^2.$$

In particular,

$$(3.5.2) \quad \text{Fix } A_T = \{x \in X \mid \|A_T(x)\| = \|x\|\}$$

and

$$(3.5.3) \quad \|A_T(x)\|^2 = \begin{cases} \|x\|^2 & \text{if } x \in M, \\ \|x\|^2 - \frac{\langle x, x - Tx \rangle^2}{\|x - Tx\|^2} & \text{if } x \notin M. \end{cases}$$

Proof. Using Lemma 3.3, we deduce that

$$\|x - y\|^2 = \|(x - A_T(x)) + (A_T(x) - y)\|^2 = \|x - A_T(x)\|^2 + \|A_T(x) - y\|^2;$$

so (3.5.1) holds. Take $y = 0$ in (3.5.1) to obtain (3.5.2). Finally, take $y = 0$ in (3.5.1) and note that $\|x - A_T(x)\|^2 = \frac{\langle x, x - Tx \rangle^2}{\|x - Tx\|^2}$ if $x \notin M$ and $\|x - A_T(x)\|^2 = 0$ if $x \in M$. This yields (3.5.3). \square

Lemma 3.6. *The following statements are equivalent:*

- (1) $Tx \in M$;
- (2) $Tx = P_M x$;
- (3) $T^n x \in M$ for every $n \geq 1$.

Proof. (1) \implies (2). If $Tx \in M$, then $Tx = P_M(Tx) = P_M x$ using Lemma 3.4(2).

(2) \implies (3). If $Tx = P_M x$, then $Tx \in M$. Thus, (3) holds when $n = 1$. We proceed by induction. If $T^n x \in M$ for some $n \geq 1$, then since $M = \text{Fix } T$, we have that

$$T^{n+1}x = T(T^n x) = T^n x \in M.$$

This completes the induction.

(3) \implies (1). Take $n = 1$. \square

The **affine hull** of a nonempty set S , denoted by $\text{aff}(S)$, is the intersection of the collection of all affine sets which contain S . (Recall that an affine set is any translation of a subspace.) Equivalently, $\text{aff}(S) = \{\alpha x + (1 - \alpha)y \mid x, y \in S, \alpha \in \mathbb{R}\}$.

Theorem 3.7. *For each $x \in X$ and $y \in M$, we have*

$$(3.7.1) \quad \|A_T(x) - y\|^2 = \|tTx + (1 - t)x - y\|^2 - (t - t_x)^2 \|Tx - x\|^2 \quad \text{for each } t \in \mathbb{R},$$

$$(3.7.2) \quad \|A_T(x) - y\| = \min_{t \in \mathbb{R}} \|tTx + (1 - t)x - y\|,$$

and the minimum is attained precisely when either $t = t_x$ if $x \notin M$ or at every $t \in \mathbb{R}$ if $x \in M$. Moreover,

$$(3.7.3) \quad d(A_T(x), M) = \min_{t \in \mathbb{R}} d(tTx + (1 - t)x, M);$$

in other words, $A_T(x)$ is the unique point in $\text{aff}\{x, Tx\}$ which is closest to M .

$$(3.7.4) \quad \|A_T(x)\| = \min_{t \in \mathbb{R}} \|tTx + (1 - t)x\|;$$

in other words, $A_T(x)$ is the unique point in $\text{aff}\{x, Tx\}$ having minimal norm. In particular,

$$(3.7.5) \quad \|A_T(x)\| \leq \min\{\|x\|, \|Tx\|\}.$$

Proof. Using Lemma 3.3, we can write

$$\begin{aligned} \|tTx + (1 - t)x - y\|^2 &= \|tTx + (1 - t)x - A_T(x) + A_T(x) - y\|^2 \\ &= \|(t - t_x)(Tx - x) + (A_T(x) - y)\|^2 \\ &= (t - t_x)^2 \|Tx - x\|^2 + \|A_T(x) - y\|^2, \end{aligned}$$

which proves (3.7.1). Equation (3.7.2) follows immediately from (3.7.1). Moreover, (3.7.3) follows by taking the infimum over all $y \in M$ in (3.7.2). Finally, (3.7.4) follows from (3.7.2) by taking $y = 0$. \square

While A_T is not linear in general, it does share some important properties of the linear mapping P_M . Namely, it is continuous, homogeneous, and “additive modulo M ”. These are recorded in parts (5), (4), and (3), respectively, of the following lemma.

Lemma 3.8. *Let $x \in X$ and $y \in M$. Then:*

- (1) $t_{x+y} = t_x$.
- (2) $t_{\alpha x} = t_x$ for every $\alpha \neq 0$.
- (3) $A_T^n(x+y) = A_T^n(x) + y$ for every $n \in \mathbb{N}$. In particular, $A_T(x+y) = A_T(x) + y$ and $A_T(y) = y$.
- (4) $A_T(\alpha x) = \alpha A_T(x)$ for every $\alpha \in \mathbb{R}$.
- (5) A_T is continuous.

Proof. (1) If $x \in M$, then $x + y \in M$ and $t_{x+y} = 1 = t_x$. If $x \notin M$, then $x + y \notin M$, and so,

$$t_{x+y} = \frac{\langle x + y, x + y - T(x + y) \rangle}{\|x + y - T(x + y)\|^2} = \frac{\langle x + y, x - Tx \rangle}{\|x - Tx\|^2} = \frac{\langle x, x - Tx \rangle}{\|x - Tx\|^2} = t_x$$

using Lemma 3.3(2).

(2) Let $\alpha \neq 0$. If $x \in M$, then $\alpha x \in M$ and $t_{\alpha x} = 1 = t_x$. If $x \notin M$, then $\alpha x \notin M$ and

$$t_{\alpha x} = \frac{\langle \alpha x, \alpha x - T(\alpha x) \rangle}{\|\alpha x - T(\alpha x)\|^2} = \frac{\langle x, x - Tx \rangle}{\|x - Tx\|^2} = t_x.$$

(3) When $n = 1$,

$$\begin{aligned} A_T(x + y) &= t_{x+y}T(x + y) + (1 - t_{x+y})(x + y) \\ &= t_x(Tx + Ty) + (1 - t_x)(x + y) \text{ using part (1)} \\ &= t_xTx + (1 - t_x)x + t_xy + (1 - t_x)y \\ &= A_T(x) + y. \end{aligned}$$

Now assume (3) holds for some $n \geq 1$. Then

$$\begin{aligned} A_T^{n+1}(x + y) &= A_T[A_T^n(x + y)] = A_T[A_T^n(x) + y] \\ &= A_T(A_T^n(x)) + y \text{ by the case } n = 1 \\ &= A_T^{n+1}(x) + y; \end{aligned}$$

so the result holds for $n + 1$.

(4) If $\alpha \neq 0$, then by (2),

$$A_T(\alpha x) = t_{\alpha x}T(\alpha x) + (1 - t_{\alpha x})(\alpha x) = t_x[\alpha T(x)] + (1 - t_x)[\alpha x] = \alpha A_T(x).$$

Since $A_T(0) = 0$, the result also holds when $\alpha = 0$.

(5) If $x \in X \setminus M$ and $x_n \rightarrow x$, then since $X \setminus M$ is open, $x_n \notin M$ eventually, and so,

$$t_{x_n} = \frac{\langle x_n, x_n - Tx_n \rangle}{\|x_n - Tx_n\|^2} \rightarrow \frac{\langle x, x - Tx \rangle}{\|x - Tx\|^2} = t_x$$

and hence A_T is continuous at x . If $x \in M$ and $\epsilon > 0$, let $y \in X$ with $\|y - x\| < \epsilon/3$. Then $\|P_M x - P_M y\| \leq \|x - y\| < \epsilon/3$ and

$$\begin{aligned} \|A_T(x) - A_T(y)\| &= \|x - A_T(y)\| \leq \|x - P_M y\| + \|P_M y - A_T(y)\| \\ &= \|x - P_M y\| + \|A_T(y - P_M y)\| \quad \text{by part (3)} \\ &\leq \|x - P_M y\| + \|y - P_M y\| \quad \text{by (3.7.5)} \\ &= \|P_M x - P_M y\| + \|y - P_M y\| \\ &< \frac{\epsilon}{3} + \|y - x\| + \|x - P_M y\| \\ &< \frac{2\epsilon}{3} + \|P_M x - P_M y\| < \epsilon. \end{aligned}$$

This proves that A_T is continuous at x . \square

Remark. We note that, while A_T is continuous, it is *not* uniformly continuous, in general, unlike a linear operator. For example, let $X = \ell_2$, let $\{e_n \mid n = 0, 1, 2, \dots\}$ be an orthonormal basis for X , and define T on X by $Tx = \sum_{n=0}^{\infty} \langle x, e_n \rangle n/(n+1) e_n$. Setting $x_n = (1/n)e_0 + ((n+1)/n)e_n$ and $y_n = e_n$ for all $n \geq 1$, we get that $\|x_n - y_n\| = (\sqrt{2}/n) \rightarrow 0$. But using the readily deduced facts that $A_T(y_n) = 0$ and $A_T(x_n) = (1/2)(e_n - e_0)$ for all n , we obtain that $\|A_T(x_n) - A_T(y_n)\| = (\sqrt{2}/2)$ for all $n \geq 1$.

Lemma 3.9. (1) $t_x \geq \frac{1}{2}$ for all $x \in X$; and
(2) $\text{Fix } A_T = M (= \text{Fix } T)$.

Proof. (1) If $x \in M$, then $t_x = 1$. If $x \notin M$, then the quadratic function,

$$q(t) := \|t(Tx - x) + x\|^2 = at^2 + bt + c,$$

where $a := \|Tx - x\|^2 > 0$, $b := 2\langle x, Tx - x \rangle$, and $c := \|x\|^2$ is strictly convex and attains its minimum at the unique point t when $q'(t) = 0$; that is, when $t = t_{\min} := -\frac{b}{2a}$. Hence,

$$t_{\min} = -\frac{2\langle x, Tx - x \rangle}{2\|Tx - x\|^2} = \frac{\langle x, x - Tx \rangle}{\|x - Tx\|^2} =: t_x.$$

But $c = q(0) = \|x\|^2$ and $\|Tx\|^2 = q(1) = a + b + c = a + b + \|x\|^2$ implies that $-b = a + \|x\|^2 - \|Tx\|^2$ and hence

$$t_x = t_{\min} = \frac{-b}{2a} = \frac{a + \|x\|^2 - \|Tx\|^2}{2a} = \frac{1}{2} + \frac{\|x\|^2 - \|Tx\|^2}{2a} \geq \frac{1}{2}.$$

(2) $x \in \text{Fix } A_T$ if and only if $x = t_x Tx + (1 - t_x)x$ if and only if $t_x(Tx - x) = 0$ if and only if $Tx - x = 0$ (using part (1)) if and only if $x \in \text{Fix } T = M$. \square

Remarks. The lower bound $\frac{1}{2}$ for t_x is sharp. To see this, take $T = -I$ and note that $t_x = \frac{1}{2}$ for every $x \in X \setminus \{0\}$. Also, if we relax the condition that T be nonexpansive and consider $T = \lambda I$ for $\lambda \neq 1$, we deduce that $t_{x,\lambda I} = \frac{1}{1-\lambda}$ for each $x \neq 0$. By varying λ , we see that $t_{x,\lambda I}$ can take on every nonzero value.

Definition 3.10. Define $f = f_T : X \rightarrow \mathbb{R}^+ := \{\alpha \in \mathbb{R} \mid \alpha \geq 0\}$ by

$$f(x) := \begin{cases} \frac{\|A_T(x) - P_M x\|}{\|Tx - P_M x\|} & \text{if } Tx \notin M, \\ 0 & \text{if } Tx \in M. \end{cases}$$

Lemma 3.11. For each $x \in X$, we have $0 \leq f(x) \leq 1$ and

$$(3.11.1) \quad \|A_T(x) - P_M x\| = f(x) \|Tx - P_M x\|.$$

Proof. This is immediate from (3.7.2) with $y = P_M x$. \square

Lemma 3.12. T commutes with P_M and P_{M^\perp} .

Proof. For each $x \in X$,

$$\begin{aligned} P_M T x &= P_M T(P_M x + P_{M^\perp} x) = P_M [T(P_M x) + T(P_{M^\perp} x)] \\ &= P_M^2 x \quad \text{since } T(M^\perp) \subset M^\perp \text{ by Lemma 3.3(4)} \\ &= P_M x = T P_M x. \end{aligned}$$

Thus, T commutes with P_M and, since $P_{M^\perp} = I - P_M$, it follows that T also commutes with P_{M^\perp} . \square

Definition 3.13. Let T be a nonexpansive linear operator on X , $M = \text{Fix } T$, and for any $n \in \mathbb{N}$, let $c_n(T)$ denote the norm of the linear operator $(T P_{M^\perp})^n$:

$$(3.13.1) \quad c_n(T) := \|(T P_{M^\perp})^n\|.$$

We will often write $c(T)$ instead of $c_1(T)$. Note that if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$, then $M := \bigcap_1^k M_i = \text{Fix } T$ and

$$(3.13.2) \quad c(T) = \|P_{M_k} P_{M_{k-1}} \cdots P_{M_1} P_{M^\perp}\| =: c(M_1, M_2, \dots, M_k)$$

is just the cosine of the angle of the k -tuple (M_1, M_2, \dots, M_k) defined by Bauschke, Borwein, and Lewis [2]. It was established in [2] that $c(T) < 1$ if and only if $M_1^\perp + M_2^\perp + \cdots + M_k^\perp$ is closed. When $k = 2$,

$$(3.13.3) \quad c(P_{M_2} P_{M_1}) = \|P_{M_2} P_{M_1} P_{M^\perp}\| = c(M_1, M_2) = c(M_2, M_1) = c(P_{M_1} P_{M_2})$$

is just the ordinary cosine of the angle between the subspaces M_1 and M_2 (see [12] or [7]).

Lemma 3.14. Let T be nonexpansive on X and $M = \text{Fix } T$. Then

(1) $c_n(T) = \|T^n - P_M\|$ for every $n \in \mathbb{N}$. In particular,

$$(3.14.1) \quad \|T^n x - P_M x\| \leq c_n(T) \|x - P_M x\| \quad \text{for every } n \in \mathbb{N}, \quad \text{and } x \in X,$$

and $c_n(T)$ is the smallest constant independent of x for which (3.14.1) is valid.

(2) $\|T^n y\| \leq c_n(T) \|y\|$ for every $y \in M^\perp$;

(3) $c_n(T) \leq c(T)^n$ for every n ;

(4) $c(T^* T) \leq c(T)^2$ and $c(T^* T) = c(T)^2$ if $\text{Fix } (T^* T) = \text{Fix } T$. In particular, if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$, then

$$(3.14.2) \quad c(T^* T) = c(T)^2;$$

(5) $\|A_T(x) - P_M x\| \leq f(x) c(T) \|x - P_M x\|$ for every $x \in X$.

Proof. (1) By Lemma 3.12, T commutes with P_M and $T P_M = P_M = P_M T$. Thus,

$$c_n(T) = \|(T P_{M^\perp})^n\| = \|[T(I - P_M)]^n\| = \|(T - P_M)^n\| = \|T^n - P_M\|.$$

Now fix any $x \in X$ and set $y = x - P_M x$. Then $y \in M^\perp$ and

$$\begin{aligned} \|T^n x - P_M x\| &= \|T^n(x - P_M x)\| = \|T^n y\| = \|T^n P_{M^\perp} y\| = \|(T P_{M^\perp})^n y\| \\ &\leq c_n(T) \|y\| = c_n(T) \|x - P_M x\|, \end{aligned}$$

which proves (3.14.1).

- (2) This was essentially proved during the course of proving (1).
- (3) $c_n(T) = \|(TP_{M^\perp})^n\| \leq \|TP_{M^\perp}\|^n = c_1(T)^n$.
- (4) Let $N = \text{Fix}(T^*T)$. Since $M = \text{Fix} T^*$ by Lemma 2.1, it follows that $M \subset N$ and so $N^\perp \subset M^\perp$. Hence, since T commutes with P_{M^\perp} by Lemma 3.12 and, by a similar proof, T^* commutes with P_{M^\perp} , we obtain

$$\begin{aligned} c(T^*T) &= \|T^*TP_{N^\perp}\| \leq \|T^*TP_{M^\perp}\| = \|(TP_{M^\perp})^*(TP_{M^\perp})\| \\ &= \|TP_{M^\perp}\|^2 = c(T)^2. \end{aligned}$$

Moreover, if $\text{Fix}(T^*T) = \text{Fix} T$, then $N = M$ and $N^\perp = M^\perp$. So the above inequality must be an equality. Equation (3.14.2) holds when T is a product of projections by Lemma 2.4.

- (5) Fix $x \in X$. Then $x - P_Mx \in M^\perp$. So Lemma 3.11 and part (1) imply
- $$\|A_T(x) - P_Mx\| = f(x)\|Tx - P_Mx\| \leq f(x)c(T)\|x - P_Mx\|.$$

□

Remark. The following example shows that the strict inequality $c(T^*T) < c(T)^2$ is possible in part (4). For let X denote the Euclidean plane and define the linear operator T on X by $Tx = x(2)e_1 + x(1)e_2$ for each $x = (x(1), x(2)) \in X$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. It is easy to verify that $\|T\| = 1$, $M := \text{Fix} T = \text{span}(e_1 + e_2)$, $c(T) = 1$, $T = T^*$, $T^*T = I$, and $c(T^*T) = c(I) = 0$.

Lemma 3.15. *Let T be nonexpansive and $M = \text{Fix} T$. Then*

- (1) *if T is normal, then $\|T^n - P_M\| = c(T)^n$ for every n ;*
 - (2) *if $\text{Fix}(T^*T) = \text{Fix} T$, then*
- $$(3.15.1) \qquad \|(T^*T)^n - P_M\| = c(T)^{2n} \quad \text{for every } n \in \mathbb{N}.$$

In particular, for every $n \in \mathbb{N}$,

$$(3.15.2) \qquad \|(P_{M_1}P_{M_2} \cdots P_{M_k}P_{M_{k-1}} \cdots P_{M_1})^n - P_M\| = c(M_1, M_2, \dots, M_k)^{2n}.$$

Proof. (1) Since T is normal and T commutes with P_{M^\perp} by Lemma 3.12, we deduce that TP_{M^\perp} is normal. Hence, using Lemma 3.14(1), we obtain

$$\|T^n - P_M\| = c_n(T) = \|(TP_{M^\perp})^n\| = \|TP_{M^\perp}\|^n = c(T)^n.$$

(2) Since T^*T is selfadjoint, hence normal, apply part (1) to T^*T instead of T using that $\text{Fix}(T^*T) = M$ to get $\|(T^*T)^n - P_M\| = c(T^*T)^n$. By Lemma 3.14(4), $c(T^*T) = c(T)^2$ and (3.15.1) follows.

(3.15.2) follows from (3.15.1) by taking $T = P_{M_k}P_{M_{k-1}} \cdots P_{M_1}$ and using Lemma 2.4 to get $\text{Fix} T^*T = \text{Fix} T$. □

The following theorem gives an upper bound on the rate of convergence of the accelerated scheme.

Theorem 3.16. *Let $x \in X$ and set*

$$x_n := A_T^{n-1}(Tx) \quad (n = 1, 2, \dots).$$

Then for every $n \in \mathbb{N}$,

$$(3.16.1) \qquad \|T^n x - P_Mx\| \leq c(T)^n \|x - P_Mx\|$$

and

$$(3.16.2) \quad \|A_T^{n-1}(Tx) - P_M x\| \leq \left[\prod_{i=1}^{n-1} f(x_i) \right] c(T)^n \|x - P_M x\|.$$

Proof. The relation (3.16.1) is a consequence of Lemma 3.14(1) and (3).

We prove (3.16.2) by induction on n . For $n = 1$, $\|Tx - P_M x\| \leq c(T)\|x - P_M x\|$ by (3.14.1). Since the product of any set of scalars over the empty set of indices is 1 by definition, (3.16.2) holds when $n = 1$. Now assume that (3.16.2) holds when $n = m \geq 1$. Then

$$\begin{aligned} \|A_T^m(Tx) - P_M x\| &= \|x_{m+1} - P_M x\| = \|A_T(x_m) - P_M x\| \\ &= \|A_T(x_m) - P_M(x_m)\| \quad (\text{by Lemma 3.4}) \\ &= f(x_m)\|T(x_m) - P_M(x_m)\| \quad (\text{by Lemma 3.11}) \\ &\leq f(x_m)c(T)\|x_m - P_M(x_m)\| \quad (\text{by (3.14.1)}) \\ &= f(x_m)c(T)\|A_T^{m-1}T(x) - P_M x\| \\ &\leq f(x_m)c(T) \left[\prod_{i=1}^{m-1} f(x_i) \right] c(T)^m \|x - P_M x\| \\ &= \left[\prod_{i=1}^m f(x_i) \right] c(T)^{m+1} \|x - P_M x\|, \end{aligned}$$

which shows that (3.16.2) holds with n replaced by $m + 1$. This completes the induction. \square

Remarks. By comparing the right sides of (3.16.1) and (3.16.2), this result seems to suggest that the accelerated algorithm is always faster than its unaccelerated counterpart by at least the factor $\left[\prod_{i=1}^{n-1} f(x_i) \right]$. Indeed, we will show below that when T is selfadjoint, nonnegative, and nonexpansive, then the accelerated method is *faster* than the original (see Theorem 3.20). In particular, the accelerated symmetric MAP is faster than the symmetric MAP. Also, the accelerated MAP for two subspaces is faster than the MAP. Perhaps surprisingly, however, we will see that this is not always the case, in general, for the accelerated MAP when there are more than two subspaces.

Theorem 3.16 can be strengthened in the particular case when $T = P_2 P_1$. To do this, it is convenient to appeal to the following simple lemma (see, e.g., [13]).

Lemma 3.17. *Let M_1 and M_2 be closed subspaces with $M = M_1 \cap M_2$ and let P_i be the orthogonal projection onto M_i for $i = 1, 2$. Then $c(P_2 P_1) = c(M_1, M_2)$ and*

- (1) *if $x \in M_1 \cap M^\perp$, then $\|P_2 x\| \leq c(M_1, M_2)\|x\|$;*
- (2) *if $x \in M_2 \cap M^\perp$, then $\|P_1 x\| \leq c(M_1, M_2)\|x\|$;*
- (3) *if $x \in M_2 \cap M^\perp$, then $\|P_2 P_1 x\| \leq c(M_1, M_2)^2\|x\|$.*

Proof. That $c(P_2 P_1) = c(M_1, M_2)$ in this case was observed following Definition 3.13.

(1) Let $x \in M_1 \cap M^\perp$. Then

$$\|P_2 x\| = \|P_2 P_1 P_{M^\perp} x\| \leq \|P_2 P_1 P_{M^\perp}\| \|x\| = c(P_2 P_1)\|x\|.$$

(2) The proof is similar to (1).

(3) Let $x \in M_2 \cap M^\perp$. Then $P_1x \in M_1 \cap M^\perp$; so by (1) and (2), we obtain

$$\|P_2P_1x\| \leq c(P_2P_1)\|P_1x\| \leq c(P_2P_1)^2\|x\|.$$

□

Theorem 3.18. *Let $T = P_{M_2}P_{M_1}$, $x \in X$, and*

$$x_n := A_T^{n-1}(Tx) \quad (n = 1, 2, \dots).$$

Then

$$(3.18.1) \qquad \|A_T^{n-1}(Tx) - P_Mx\| \leq \left[\prod_1^{n-1} f(x_i) \right] c(M_1, M_2)^{2n-1} \|x - P_Mx\|.$$

Proof. The proof is by induction and proceeds just as in the proof of Theorem 3.16. The only point that should be noted is that in the induction step, we use the inequality $\|T(x_m) - P_M(x_m)\| \leq c(T)^2\|x_m - P_M(x_m)\|$ (rather than the same expression with $c(T)$ instead of $c(T)^2$ that was used in Theorem 3.16). The proof of this inequality follows immediately from Lemma 3.17(3). □

Remarks. (1) Gearhart and Koshy [13] established (a weaker version of) the special case of Theorem 3.18 when $c := c(M_1, M_2) < 1$ and with an additional factor ρ on the right side of (3.18.1), where $\rho := \frac{1}{\sqrt{1-c^2}} \geq 1$.

(2) The inequality (3.18.1) improves the bound on the ordinary MAP in case $k = 2$, due to Aronszajn [1], who showed that

$$(3.18.2) \qquad \|(P_2P_1)^n x - P_Mx\| \leq c(M_1, M_2)^{2n-1} \|x - P_Mx\| \quad \text{for all } x \in X.$$

In fact, Kayalar and Weinert [17] showed that the Aronszajn bound is *sharp*, i.e., $\|(P_2P_1)^n - P_M\| = c(M_1, M_2)^{2n-1}$.

Next we show that the accelerated algorithms are always at least as fast as their unaccelerated counterparts provided that T is selfadjoint, nonnegative, and nonexpansive. It is first convenient to establish the following result.

Lemma 3.19. *If*

$$(3.19.1) \qquad \|T^{n-1}(A_T(x))\| \leq \|T^n x\| \quad \text{for every } x \in M^\perp \text{ and } n \in \mathbb{N},$$

then

$$(3.19.2) \qquad \|A_T^{n-1}(Tx)\| \leq \|T^n x\| \quad \text{for every } x \in M^\perp \text{ and } n \in \mathbb{N}.$$

In particular, if (3.19.1) holds and the original algorithm converges, then

$$(3.19.3) \qquad \|A_T^{n-1}(Tx) - P_Mx\| \leq \|T^n x - P_Mx\| \quad \text{for every } x \in X, n \in \mathbb{N},$$

and hence the accelerated algorithm converges at least as fast as the original.

Proof. When $n = 1$, (3.19.2) is trivial. If $n \geq 2$, then for each $x \in M^\perp$,

$$\begin{aligned} \|A_T^{n-1}(Tx)\| &= \|A_T(A_T^{n-2}(Tx))\| \leq \|T(A_T^{n-2}(Tx))\| \quad \text{using (3.7.5)} \\ &= \|T(A_T(y))\|, \quad \text{where } y := A_T^{n-3}(Tx) \in M^\perp \text{ by Lemma 3.3(4)} \\ &\leq \|T^2y\| \quad \text{by (3.19.1)} \\ &= \|T^2(A_T^{n-3}(Tx))\| \\ &= \|T^2(A_T(z))\|, \quad \text{where } z := A_T^{n-4}(Tx) \in M^\perp \text{ by Lemma 3.3(4)} \\ &\leq \|T^3z\| \quad \text{by (3.19.1)} \\ &= \|T^3(A_T^{n-4}(Tx))\|. \end{aligned}$$

Continuing in this way, we end up with the inequality $\|A_T^{n-1}(Tx)\| \leq \|T^n x\|$, which verifies (3.19.2) when $n \geq 2$.

To verify the last statement, let $x \in X$. Then $x - P_M x \in M^\perp$ and so by (3.19.2) and Lemma 3.8(3), we get

$$\|A_T^{n-1}(Tx) - P_M x\| = \|A_T^{n-1}(T(x - P_M x))\| \leq \|T^n(x - P_M x)\| = \|T^n x - P_M x\|$$

and this verifies (3.19.3). \square

The natural question raised by Lemma 3.19 is this: for which T does (3.19.1) hold? We will show next that if T is selfadjoint, nonnegative, and nonexpansive, then (3.19.1) and hence (3.19.3) hold. It should be noted that our proof seems to use the spectral theorem (for compact selfadjoint operators) in an essential way.

Theorem 3.20. *Let T be selfadjoint, nonnegative, and nonexpansive. Then*

$$(3.20.1) \quad \|A_T^{n-1}(Tx) - P_M x\| \leq \|T^n x - P_M x\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}.$$

In other words, the accelerated algorithm converges at least as fast as its unaccelerated counterpart.

Corollary 3.21. *If $T = P_1 P_2 \cdots P_k P_{k-1} \cdots P_1$, then*

$$(3.21.1) \quad \|A_T^{n-1}(Tx) - P_M x\| \leq \|T^n x - P_M x\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}.$$

In other words, the accelerated symmetric MAP is at least as fast as the symmetric MAP.

The corollary follows since $T = Q^*Q$, where $Q = P_k P_{k-1} \cdots P_1$.

Proof of Theorem 3.20. By Lemma 3.19, it suffices to show that

$$(3.20.2) \quad \|T^{m-1}A_T(y)\| \leq \|T^m y\| \quad \text{for every } y \in M^\perp \text{ and } m \in \mathbb{N}.$$

Toward this end, fix $y \in M^\perp$ and $m \in \mathbb{N}$. If $y = 0$, (3.20.2) is trivial. Thus, by scaling and Lemma 3.8(4), we may assume $\|y\| = 1$. If $m = 1$, then (3.20.2) follows from (3.7.5). Thus, we may assume $m \geq 2$. Let

$$N = \text{span}\{y, Ty, T^2y, \dots, T^m y\}.$$

By Lemma 3.3(4), $N \subset M^\perp$. Define $S := P_N T P_N$. Then S is compact, selfadjoint, nonexpansive, $\mathcal{R}(S) := \text{Range of } S \subset N$, and so $n := \dim \mathcal{R}(S) \leq m + 1$. We may assume that $Ty \neq 0$. For if $Ty = 0$, then $A_T(y) = 0$ by (3.7.5); so (3.20.2) holds and we are done. But if $Ty \neq 0$, then $Sy \neq 0$ and hence $n \geq 1$. As a consequence of

the Spectral Theorem [3, Corollary 5.4, p. 47], we readily deduce that there exists an orthonormal set of n eigenvectors $\{v_1, v_2, \dots, v_n\}$ of S such that

$$(3.20.3) \quad Sx := \sum_1^n \lambda_i \langle x, v_i \rangle v_i \quad \text{for every } x \in X,$$

where λ_i is the (nonzero) eigenvalue corresponding to $v_i : Sv_i = \lambda_i v_i$ ($i = 1, 2, \dots, n$). In particular, $\{v_1, \dots, v_n\}$ is an orthonormal basis for $\mathcal{R}(S)$. Since T is nonnegative,

$$\begin{aligned} \lambda_i &= \lambda_i \langle v_i, v_i \rangle = \langle \lambda_i v_i, v_i \rangle = \langle Sv_i, v_i \rangle = \langle P_N T P_N v_i, v_i \rangle \\ &= \langle T P_N v_i, P_N v_i \rangle = \langle T v_i, v_i \rangle \quad \text{since } v_i \in N \\ &\geq 0. \end{aligned}$$

Thus, $\lambda_i > 0$ for each i . Since S is nonexpansive,

$$\lambda_i = \|\lambda_i v_i\| = \|Sv_i\| \leq \|v_i\| = 1.$$

We have shown that $0 < \lambda_i \leq 1$ for each i . Moreover, if some $\lambda_i = 1$, then

$$\begin{aligned} 1 &= \langle v_i, v_i \rangle = \langle v_i, Sv_i \rangle = \langle v_i, P_N T P_N v_i \rangle \\ &= \langle v_i, T P_N v_i \rangle = \langle v_i, T v_i \rangle \leq \|v_i\| \|T v_i\| \leq 1. \end{aligned}$$

So equality must hold throughout this string of inequalities. Using the condition of equality in Schwarz's inequality, we obtain $Tv_i = \rho v_i$ for some $\rho > 0$ and $\|Tv_i\| = \|v_i\| = 1$. Hence, $\rho = 1$ and $Tv_i = v_i$. That is, $v_i \in \text{Fix } T = M$. But $v_i \in M^\perp$ implies that $v_i = 0$, a contradiction. This proves that $\lambda_i < 1$ for each i . Hence, we have shown that

$$(3.20.4) \quad 0 < \lambda_i < 1 \quad \text{for } i = 1, 2, \dots, n.$$

Let $\alpha_i := \langle y, v_i \rangle$ for each i .

Claim 1. $T^j y = S^j y = \sum_{i=1}^n \alpha_i \lambda_i^j v_i \quad (j = 1, 2, \dots, m)$.

The formula for S , $S^j y = \sum_1^n \alpha_i \lambda_i^j v_i$, follows easily from (3.20.3) and the fact that $Sv_i = \lambda_i v_i$. To prove the corresponding statement about T , we proceed by induction on j . For $j = 1$, since y and Ty are in N , we obtain $Ty = P_N Ty = P_N T P_N y = Sy$; so the result holds when $j = 1$. Now suppose the result holds when $j = l \leq m - 1$. Then

$$S^{l+1} y = S(S^l y) = S(T^l y) = P_N T P_N (T^l y) = P_N T (T^l y) = P_N T^{l+1} y = T^{l+1} y$$

since $T^{l+1} y \in N$. This proves the claim.

Since $\mathcal{R}(S)^\perp = \mathcal{N}(S^*) = \mathcal{N}(S)$, where $\mathcal{N}(S)$ is the null space of S , we have that $X = \mathcal{R}(S) \oplus \mathcal{N}(S)$ and hence we can write y as $y = y_1 + y_0$, where $y_1 \in \mathcal{R}(S) = \text{span}\{v_1, v_2, \dots, v_n\}$ and $y_0 \in \text{span}\{v_1, v_2, \dots, v_n\}^\perp = \mathcal{N}(S)$. Then

$$y = \sum_1^n \langle y_1, v_i \rangle v_i + y_0 = \sum_1^n \langle y, v_i \rangle v_i + y_0 = \sum_1^n \alpha_i v_i + y_0$$

and

$$\|y\|^2 = \sum_1^n \alpha_i^2 + \|y_0\|^2.$$

Claim 2. $T^{m-1} A_T(y) = \sum_{i=1}^n \alpha_i \lambda_i^{m-1} \{1 - (1 - \lambda_i) t_y\} v_i$.

We compute

$$\begin{aligned} T^{m-1}A_T(y) &= T^{m-1}[t_yTy + (1-t_y)y] = t_yT^my + (1-t_y)T^{m-1}y \\ &= t_yS^my + (1-t_y)S^{m-1}y = t_y\sum_1^n \alpha_i\lambda_i^m v_i + (1-t_y)\sum_1^n \alpha_i\lambda_i^{m-1}v_i \\ &= \sum_1^n \alpha_i\lambda_i^{m-1}\{t_y\lambda_i + (1-t_y)\}v_i = \sum_1^n \alpha_i\lambda_i^{m-1}\{1 - (1-\lambda_i)t_y\}v_i \end{aligned}$$

which proves the claim.

By Claims 1 and 2, we see that (3.20.2) holds if and only if

$$\sum_{i=1}^n \alpha_i^2 \lambda_i^{2m-2} \{1 - (1-\lambda_i)t_y\}^2 \leq \sum_{i=1}^n \alpha_i^2 \lambda_i^{2m}$$

which, after some algebra, may be rewritten as

$$(3.20.5) \quad q(t_y) \leq 0,$$

where

$$(3.20.6) \quad \begin{aligned} q(t) &:= \alpha t^2 - 2\beta t + \gamma, & \alpha &:= \sum_1^n \alpha_i^2 \lambda_i^{2m-2} (1-\lambda_i)^2, \\ \beta &:= \sum_1^n \alpha_i^2 \lambda_i^{2m-2} (1-\lambda_i), & \gamma &:= \sum_1^n \alpha_i^2 \lambda_i^{2m-2} (1-\lambda_i^2). \end{aligned}$$

Claim 3. The function h , defined on the nonnegative real line by

$$h(t) := \frac{\sum_i \alpha_i^2 \lambda_i^t (1-\lambda_i)}{\sum_j \alpha_j^2 \lambda_j^t (1-\lambda_j)^2} \quad \text{for all } t \geq 0,$$

is increasing.

Writing $h(t) = u(t)/v(t)$, it suffices to verify that $h'(t) \geq 0$. Equivalently, it suffices to show that

$$(3.20.7) \quad u'(t)v(t) \geq u(t)v'(t) \quad \text{for all } t \geq 0.$$

Setting

$$\beta_i = \frac{\alpha_i^2 (1-\lambda_i) \lambda_i^t}{\sum_j \alpha_j^2 (1-\lambda_j) \lambda_j^t},$$

we see that $\beta_i \geq 0$, $\sum_1^n \beta_i = 1$, and (3.20.7) may be rewritten as

$$(3.20.8) \quad \sum_j \beta_j \lambda_j \ln \lambda_j \geq \left(\sum_i \beta_i \ln \lambda_i \right) \left(\sum_j \beta_j \lambda_j \right).$$

Since the function $t \mapsto t \ln t$ is convex on $(0, \infty)$, it follows that

$$(3.20.9) \quad \left(\sum_j \beta_j \lambda_j \right) \ln \left(\sum_i \beta_i \lambda_i \right) \leq \sum_j \beta_j \lambda_j \ln \lambda_j.$$

On the other hand, the function $t \mapsto \ln t$ is concave on $(0, \infty)$; so

$$(3.20.10) \quad \ln \left(\sum_j \beta_j \lambda_j \right) \geq \sum_j \beta_j \ln \lambda_j.$$

Combining (3.20.9) and (3.20.10), we obtain (3.20.8), and this proves Claim 3.

To prove (3.20.5), and finish the proof of the theorem, we must verify that $q(t_y) \leq 0$, where q is the quadratic defined in (3.20.6). Now $q(0) = \gamma > 0$ and $q(1) = \alpha - 2\beta + \gamma = 0$. Also, an inspection of the coefficients shows that $0 < \alpha < \beta < \gamma$. Further, the quadratic formula shows that the zeros of q are given by

$$t_{\min} = \frac{\beta - \sqrt{\beta^2 - \alpha\gamma}}{\alpha}, \quad \text{and } t_{\max} = \frac{\beta + \sqrt{\beta^2 - \alpha\gamma}}{\alpha}.$$

Since $\beta = \frac{1}{2}(\alpha + \gamma)$, it follows that $t_{\min} = 1$ and $t_{\max} = \gamma/\alpha > 1$. Since q has a positive leading coefficient, we see that $q(t) \leq 0$ if and only if $t_{\min} \leq t \leq t_{\max}$, i.e., $1 \leq t \leq \gamma/\alpha$. Thus to prove $q(t_y) \leq 0$, we must show that

$$(3.20.11) \quad 1 \leq t_y \leq \frac{\gamma}{\alpha}.$$

We have, using Claim 1, that

$$\begin{aligned} t_y &= \frac{\langle y, y - Ty \rangle}{\|y - Ty\|^2} = \frac{\langle \sum_i \alpha_i v_i + y_0, \sum_i \alpha_i (1 - \lambda_i) v_i + y_0 \rangle}{\sum_i \alpha_i^2 (1 - \lambda_i)^2 + \|y_0\|^2} \\ &= \frac{\sum_i \alpha_i^2 (1 - \lambda_i) + \|y_0\|^2}{\sum_i \alpha_i^2 (1 - \lambda_i)^2 + \|y_0\|^2}. \end{aligned}$$

Since $0 < (1 - \lambda_i)^2 < 1 - \lambda_i$, it follows that $t_y \geq 1$. Also, $t_y \leq \gamma/\alpha$ is equivalent to

$$(3.20.12) \quad \frac{\sum_i \alpha_i^2 (1 - \lambda_i) + \|y_0\|^2}{\sum_j \alpha_j^2 (1 - \lambda_j)^2 + \|y_0\|^2} \leq \frac{\sum_i \alpha_i^2 \lambda_i^{2m-2} (1 - \lambda_i)}{\sum_j \alpha_j^2 \lambda_j^{2m-2} (1 - \lambda_j)^2}.$$

But

$$(3.20.13) \quad \frac{\sum_i \alpha_i^2 (1 - \lambda_i) + \|y_0\|^2}{\sum_j \alpha_j^2 (1 - \lambda_j)^2 + \|y_0\|^2} \leq \frac{\sum_i \alpha_i^2 (1 - \lambda_i)}{\sum_j \alpha_j^2 (1 - \lambda_j)^2}$$

follows since $\sum_i \alpha_i^2 (1 - \lambda_i) \geq \sum_i \alpha_i^2 (1 - \lambda_i)^2$.

By Claim 3, h is increasing so that $h(0) \leq h(2m - 2)$. That is,

$$(3.20.14) \quad \frac{\sum_i \alpha_i^2 (1 - \lambda_i)}{\sum_j \alpha_j^2 (1 - \lambda_j)^2} \leq \frac{\sum_i \alpha_i^2 \lambda_i^{2m-2} (1 - \lambda_i)}{\sum_j \alpha_j^2 \lambda_j^{2m-2} (1 - \lambda_j)^2}.$$

Combining (3.20.13) and (3.20.14), we obtain (3.20.12) and hence $t_y \leq \gamma/\alpha$. This proves (3.20.12), and completes the proof of the theorem. \square

A certain analogue of Theorem 3.20, valid when T is not selfadjoint, can be deduced from Theorem 3.20 as follows.

Corollary 3.22. *Suppose S is a bounded linear operator on X , L is a closed subspace of X such that $L \supset \mathcal{R}(S)$, and SP_L is selfadjoint, nonnegative, and non-expansive. Let $M = \text{Fix } S$. Then*

$$(3.22.1) \quad \|A_S^{n-1} SP_L x - P_M x\| \leq \|S^n P_L x - P_M x\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}.$$

In particular,

$$(3.22.2) \quad \|A_S^{n-1} Sx - P_M x\| \leq \|S^n x - P_M x\| \quad \text{for each } x \in L \text{ and } n \in \mathbb{N}.$$

Proof. Set $T = SP_L$. Then T satisfies the hypothesis of Theorem 3.20. Moreover, since $\mathcal{R}(S) \subset L$, it follows that $\text{Fix } T = \text{Fix } S = M$. Thus, we deduce from (3.20.1) that

$$(3.22.3) \quad \|A_T^{n-1}(Tx) - P_M x\| \leq \|T^n x - P_M x\| \quad \text{for each } x \in X \text{ and } n \in \mathbb{N}.$$

Since $\mathcal{R}(S) \subset L$, we deduce that

$$T^n = (SP_L)^n = S(P_LS)^{n-1}P_L = S(S)^{n-1}P_L = S^n P_L.$$

In particular, $T^n x = S^n x$ for each $x \in L$. Moreover, for each $y \in L$,

$$A_T(y) = t_{y,T}Ty + (1 - t_{y,T})y = t_{y,T}Sy + (1 - t_{y,T})y$$

and $A_Sy = t_{y,S}Sy + (1 - t_{y,S})y$. But

$$t_{y,T} = \frac{\langle y, y - Ty \rangle}{\|y - Ty\|^2} = \frac{\langle y, y - Sy \rangle}{\|y - Sy\|^2} = t_{y,S};$$

so $A_T(y) = A_Sy \in L$ so that, inductively, $A_T^{n-1}(y) = A_S^{n-1}y$. Substituting back into (3.22.3), we obtain (3.22.2). In general, for any $y \in X$, $x = P_Ly \in L$, and so

$$(3.22.4) \quad \|A_S^{n-1}SP_Ly - P_MP_Ly\| \leq \|S^n P_Ly - P_MP_Ly\|.$$

But $M \subset \mathcal{R}(S) \subset L$; so $P_MP_Ly = P_My$ and substituting this into (3.22.4) yields (3.22.1). \square

One application of Corollary 3.22 is in the case of the MAP for two subspaces.

Theorem 3.23. *Let M_1 and M_2 be closed subspaces in X , $Q = P_2P_1$, and $M = M_1 \cap M_2$. Then for each $n \in \mathbb{N}$,*

$$(3.23.1) \quad \|A_Q^{n-1}Qx - P_Mx\| \leq \|Q^n x - P_Mx\| \quad \text{for every } x \in X.$$

In other words, the accelerated MAP is faster than the MAP in the case of two subspaces.

Proof. Take $S = Q$ and $L = M_2$ in Corollary 3.22 to obtain

$$(3.23.2) \quad \|A_Q^{n-1}QP_2x - P_Mx\| \leq \|Q^n P_2x - P_Mx\| \quad \text{for every } x \in X.$$

In particular, (3.23.1) holds for each $x \in M_2$. It remains to show that (3.23.1) holds for all $x \in X$. We first verify

$$(3.23.3) \quad \overline{\mathcal{R}(P_2P_1P_2)} = \overline{\mathcal{R}(P_2P_1)}.$$

To see this, note that it is well-known that for any bounded linear operator T on X ,

$$(3.23.4) \quad \mathcal{N}(T^*T) = \mathcal{N}(T) \quad \text{and} \quad \mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}.$$

Putting $T = P_1P_2$ in (3.23.4), we obtain that $\mathcal{N}(P_2P_1P_2) = \mathcal{N}(P_1P_2)$ and hence $\overline{\mathcal{R}(P_2P_1P_2)} = \mathcal{N}(P_2P_1P_2)^\perp = \mathcal{N}(P_1P_2)^\perp = \overline{\mathcal{R}(P_2P_1)}$, which proves (3.23.3).

Now fix any $x \in X$ and set $z = P_2P_1x$. Then $z \in \mathcal{R}(P_2P_1)$ and so, by (3.23.3), we obtain that $z = \lim z_k$, where $z_k \in \mathcal{R}(P_2P_1P_2)$ for each k . Then we choose $w_k \in X$ so that $z_k = P_2P_1P_2w_k$. Let $y_k := P_2w_k$. Then $y_k \in M_2$ and $z_k = P_2P_1y_k$. Since P_i commutes with P_M for $i = 1, 2$, we have that

$$(3.23.5) \quad \begin{aligned} P_Mx &= P_2P_1P_Mx = P_MP_2P_1x = P_Mz = \lim_k P_Mz_k \\ &= \lim_k P_MP_2P_1y_k = \lim_k P_2P_1P_My_k = \lim_k P_My_k. \end{aligned}$$

Moreover,

$$(3.23.6) \quad \lim_k Qy_k = \lim_k P_2P_1y_k = \lim_k z_k = z = Qx.$$

By (3.23.2) applied to $y_k \in M_2$, we obtain that

$$(3.23.7) \quad \|A_Q^{n-1}Qy_k - P_My_k\| \leq \|Q^n y_k - P_My_k\|.$$

Letting $k \rightarrow \infty$ in (3.23.7), and using (3.23.5), (3.23.6), and the continuity of A_Q (Lemma 3.8(5)), we obtain (3.23.1). \square

The following is an example showing that the accelerated MAP may be *slower* than the MAP when there are more than two subspaces!

Example 3.24. Let $X = \ell_2$ and let e_i ($i = 1, 2, \dots$) denote the canonical unit vectors in X : $e_i(j) = \delta_{ij}$ for all i, j . Define five 2-dimensional subspaces as follows:

$$\begin{aligned} M_1 &= \text{span}\{e_2, e_3\}, & M_2 &= \text{span}\{e_2 + e_4, e_3 + e_5\}, & M_3 &= \text{span}\{e_4, e_5\}, \\ M_4 &= \text{span}\{e_1 + e_4, e_2 + e_5\}, & \text{and} & & M_5 &= \text{span}\{e_1, e_2\}. \end{aligned}$$

Let $P_i = P_{M_i}$ for $i = 1, 2, \dots, 5$ and $T := P_5 P_4 P_3 P_2 P_1$. It is easy to verify that

$$Tx = \frac{1}{4}x(2)e_1 + \frac{1}{4}x(3)e_2 \quad \text{for each } x \in \ell_2.$$

Also, $\|T\| = \frac{1}{4}$ and $M := \text{Fix } T = \{0\}$. Set $x_0 := 4e_3$. Then $Tx_0 = e_2$, $T^2x_0 = \frac{1}{4}e_1$, and $T^n x_0 = 0$ for all $n \geq 3$.

Let $z_0 := Tx_0 = e_2$ and define $z_n := A_T(z_{n-1}) = A_T^n(z_0)$ for $n \geq 1$. Since the range of T is $\text{span}\{e_1, e_2\}$ and $A_T(x)$ is an affine combination of Tx and x , it follows that

$$(3.24.1) \quad z_n = \alpha_n e_1 + \beta_n e_2 \quad (n = 0, 1, \dots)$$

for some scalars α_n, β_n . We will prove that $z_n \neq 0$ for every n .

Having done this, we would then obtain for every $n \geq 3$ that

$$\|A_T^{n-1}(Tx_0) - P_M x_0\| = \|A_T^{n-1}(Tx_0)\| = \|z_{n-1}\| > 0 = \|T^n x_0\| = \|T^n x_0 - P_M x_0\|$$

which shows that the accelerated MAP is *slower* than the MAP beginning with the third iterate. (It should be noted, however, that the second iterate for the accelerated method has a strictly smaller norm than the corresponding unaccelerated term: $\|A_T(Tx_0)\| = 1/\sqrt{17} < 1/4 = \|T^2x_0\|$.)

It remains to show that $z_n \neq 0$ for each n , and this will be done through a series of claims.

Claim 1. $Tz_n = \frac{1}{4}\beta_n e_1$ ($n = 0, 1, \dots$).

This follows from

$$Tz_n = T(\alpha_n e_1 + \beta_n e_2) = \alpha_n T e_1 + \beta_n T e_2 = \frac{1}{4}\beta_n e_1.$$

Next we prove

Claim 2. $z_{n+1} = 0$ if and only if $\beta_n = 0$.

For suppose $z_{n+1} = 0$. Then

$$\begin{aligned} 0 &= A_T(z_n) = t_n Tz_n + (1 - t_n)z_n, \quad \text{where } t_n = t_{z_n} \\ &= \frac{1}{4}\beta_n t_n e_1 + (1 - t_n)(\alpha_n e_1 + \beta_n e_2) \\ &= \left[\frac{1}{4}\beta_n t_n + (1 - t_n)\alpha_n\right] e_1 + (1 - t_n)\beta_n e_2. \end{aligned}$$

It follows that

$$\frac{1}{4}\beta_n t_n + (1 - t_n)\alpha_n = 0 \quad \text{and} \quad (1 - t_n)\beta_n = 0.$$

No matter what the value of t_n is, the two equations above imply $\beta_n = 0$.

Conversely, if $\beta_n = 0$, then $z_n = \alpha_n e_1$ implies that $Tz_n = 0$ and hence $A_T(z_n) = 0$ (since $\|A_T(z_n)\| \leq \|Tz_n\|$ by (3.7.5)). Thus, $z_{n+1} = A_T(z_n) = 0$.

Claim 3. For $n = 0, 1, 2, \dots$,

$$(3.24.2) \quad \beta_{n+1}\beta_n = \alpha_{n+1} \left(\frac{1}{4}\beta_n - \alpha_n \right).$$

In particular, if $\beta_n \neq 0$, then

$$(3.24.3) \quad \beta_{n+1} = \alpha_{n+1} \left(\frac{1}{4} - \frac{\alpha_n}{\beta_n} \right).$$

To verify this, note that by Lemma 3.3(2), we obtain that

$$\langle z_{n+1}, z_n - Tz_n \rangle = \langle A_T(z_n), z_n - Tz_n \rangle = 0.$$

Using the representation of z_n in (3.24.1), we expand the above equation and deduce that $\alpha_{n+1}(\alpha_n - \frac{1}{4}\beta_n) + \beta_{n+1}\beta_n = 0$, which is just (3.24.2).

If the result that $z_n \neq 0$ for each n were false, we let n_0 denote the *smallest* integer such that $z_{n_0+1} = 0$. Now $\beta_0 = 1$ and one can readily compute that

$$z_1 = A_T(z_0) = A_T(e_2) = t_{e_2}Te_2 + (1 - t_{e_2})e_2 = \frac{1}{4}t_{e_2}e_1 + (1 - t_{e_2})e_2,$$

where

$$t_{e_2} = \frac{\langle e_2, e_2 - Te_2 \rangle}{\|e_2 - Te_2\|^2} = \frac{1}{\|e_2 - \frac{1}{4}e_1\|^2} = \frac{16}{17}.$$

Hence, $z_1 = \frac{4}{17}e_1 + \frac{1}{17}e_2$ and so $\alpha_1 = \frac{4}{17}$ and $\beta_1 = \frac{1}{17}$. Thus $\beta_0 \neq 0$ and $\beta_1 \neq 0$. By Claim 2, $\beta_{n_0} = 0$; so $n_0 \geq 2$, and $\beta_n \neq 0$ for every $n \leq n_0 - 1$. Further, by Claim 3, we deduce

$$(3.24.4) \quad \beta_{n+1} = \alpha_{n+1} \left(\frac{1}{4} - \mu_n \right) \quad \text{for } n = 0, 1, \dots, n_0 - 1,$$

where $\mu_n := \alpha_n/\beta_n$.

From (3.24.4), we deduce that $\alpha_{n+1} \neq 0$ whenever $\beta_{n+1} \neq 0$ and $0 \leq n \leq n_0 - 1$. Since $\beta_{k+1} \neq 0$ for $0 \leq k \leq n_0 - 2$, it follows that $\alpha_{k+1} \neq 0$ for $0 \leq k \leq n_0 - 2$. In other words,

$$(3.24.5) \quad \alpha_n \neq 0 \quad \text{and} \quad \beta_n \neq 0 \quad \text{for } 1 \leq n \leq n_0 - 1.$$

Using (3.24.4), we obtain that

$$(3.24.6) \quad 0 \neq \frac{\beta_{n+1}}{\alpha_{n+1}} = \frac{1}{4} - \mu_n \quad \text{for } 0 \leq n \leq n_0 - 2.$$

Next consider the following subset of the rational numbers:

$$\mathbb{Q}^* := \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}, p \text{ even}, q \text{ odd} \right\}.$$

In particular, $0 \in \mathbb{Q}^*$ but $\frac{1}{4} \notin \mathbb{Q}^*$.

Claim 4. The function $f(x) = (\frac{1}{4} - x)^{-1}$ maps \mathbb{Q}^* into \mathbb{Q}^* .

First, note that f is well-defined since $\frac{1}{4} \notin \mathbb{Q}^*$. Next, let $x \in \mathbb{Q}^*$. Then $x = \frac{p}{q}$ for some even p and odd q . Hence,

$$f(x) = \frac{1}{\frac{1}{4} - \frac{p}{q}} = \frac{4q}{q - 4p}.$$

Since $4q$ is even and $q - 4p$ is odd, it follows that $f(x) \in \mathbb{Q}^*$.

Claim 5. $\mu_n \in \mathbb{Q}^*$ for $0 \leq n \leq n_0 - 1$. In particular, $\mu_n \neq \frac{1}{4}$ for $0 \leq n \leq n_0 - 1$.

To verify this, first note that $\mu_0 = \frac{\alpha_0}{\beta_0} = 0 \in \mathbb{Q}^*$. By (3.24.6), it follows that

$$(3.24.7) \quad \mu_{n+1} := \frac{\alpha_{n+1}}{\beta_{n+1}} = \frac{1}{\frac{1}{4} - \mu_n} \quad (n = 0, 1, \dots, n_0 - 2).$$

Using (3.24.7), Claim 4, and induction, it follows that $\mu_{n+1} \in \mathbb{Q}^*$ for $n = 0, 1, \dots, n_0 - 2$. This proves Claim 5.

Finally, $\mu_{n_0-1} \neq \frac{1}{4}$ from Claim 5. Since $\beta_{n_0} = 0$, (3.24.4) implies that $\alpha_{n_0} = 0$. But then $z_{n_0} = \alpha_{n_0}e_1 + \beta_{n_0}e_2 = 0$, which contradicts the choice of n_0 . This proves that the accelerated MAP is *slower* than the MAP for this example. However, both the MAP and the accelerated MAP do converge! This raises an interesting question that we pose now.

Open Problem. Let T be a nonexpansive mapping on X which is asymptotically regular, and let $M = \text{Fix } T$. Then, by Corollary 2.3, the algorithm converges:

$$(3.24.8) \quad \lim_{n \rightarrow \infty} \|T^n x - P_M x\| = 0 \quad \text{for each } x \in X.$$

Is it true that the *accelerated algorithm* for T also converges? That is, does the following hold:

$$(3.24.9) \quad \lim_{n \rightarrow \infty} \|A_T^n(Tx) - P_M x\| = 0 \quad \text{for each } x \in X?$$

We have seen that the answer is *affirmative* in several special cases. For example, when any one of the following conditions are satisfied, then (3.24.9) holds.

- (1) T is selfadjoint and nonnegative (Theorem 3.20); in particular, if $T = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^* (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})$ (Corollary 3.21).
- (2) $T = P_{M_2} P_{M_1}$ is the product of two orthogonal projections (Theorem 3.23).
- (3) $c(T) < 1$ (Theorem 3.16); in particular, if $T = P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$ and $M_1^\perp + M_2^\perp + \cdots + M_k^\perp$ is closed, then $c(T) < 1$ (see [2]).

In particular, does (3.24.9) hold if T is the product of $k \geq 3$ orthogonal projections? In this case, we *can* show that

$$(3.24.10) \quad A_T^n(Tx) \rightarrow P_M x \quad \text{weakly for each } x \in X.$$

But we are not sure whether the convergence must be in norm.

To prepare for the last main result, we begin with a useful lemma.

Lemma 3.25. Define the function

$$E(\alpha, \beta) := \frac{\beta - \alpha}{2 - \alpha - \beta} \quad \text{for all } \alpha, \beta \in \mathbb{R} \text{ with } \alpha + \beta \neq 2.$$

- (1) Then E is a continuously differentiable function on its domain such that

- (a) $\frac{\partial E(\alpha, \beta)}{\partial \alpha} = \frac{2(\beta - 1)}{(2 - \alpha - \beta)^2}$, and
- (b) $\frac{\partial E(\alpha, \beta)}{\partial \beta} = \frac{2(1 - \alpha)}{(2 - \alpha - \beta)^2}$.

In particular, if $c \leq 1$, then $E(\alpha, c)$ (respectively, $E(c, \beta)$) is a decreasing (respectively, increasing) function of α (respectively, β) in each of the two components of its domain.

- (2) (a) $|E(\alpha, \beta)| < 1$ if and only if $(1 - \alpha)(1 - \beta) > 0$.
- (b) $|E(\alpha, \beta)| = 1$ if and only if $(1 - \alpha)(1 - \beta) = 0$.
- (c) $|E(\alpha, \beta)| > 1$ if and only if $(1 - \alpha)(1 - \beta) < 0$.

Proof. The verification of (1) is easy.

(2) Write

$$E(\alpha, \beta) = \frac{\beta - \alpha}{2 - \alpha - \beta} = \frac{(1 - \alpha) - (1 - \beta)}{(1 - \alpha) + (1 - \beta)} = \frac{r_1 - r_2}{r_1 + r_2},$$

where $r_1 = 1 - \alpha$, and $r_2 = 1 - \beta$. Clearly, $|E(\alpha, \beta)| < 1$ if and only if $|\frac{r_1 - r_2}{r_1 + r_2}| < 1$ if and only if $|r_1 - r_2| < |r_1 + r_2|$ if and only if $r_1 r_2 > 0$. This proves (a). Also, $|E(\alpha, \beta)| = 1$ if and only if $r_1 r_2 = 0$, which proves (b). Finally, $|E(\alpha, \beta)| > 1$ if and only if $|\frac{r_1 - r_2}{r_1 + r_2}| > 1$ if and only if $r_1 r_2 < 0$, which proves (c). \square

Lemma 3.26. *Let T be selfadjoint,*

$$(3.26.1) \quad c_1 := \inf\{\langle Tx, x \rangle \mid x \in M^\perp, \|x\| = 1\},$$

and

$$(3.26.2) \quad c_2 := \sup\{\langle Tx, x \rangle \mid x \in M^\perp, \|x\| = 1\},$$

where both c_1 and c_2 are defined to be 0 if $M^\perp = \{0\}$, i.e., if $M = X$. Then

$$(3.26.3) \quad \max\{c_2, -c_1\} = c(T) := \|TP_{M^\perp}\|.$$

Moreover, if T is also nonnegative, then

$$(3.26.4) \quad c_2 = c(T).$$

Proof. First note that

$$-c_1 = -\inf\{\langle Tx, x \rangle \mid x \in M^\perp, \|x\| = 1\} = \sup\{-\langle Tx, x \rangle \mid x \in M^\perp, \|x\| = 1\}.$$

Hence,

$$\begin{aligned} \max\{c_2, -c_1\} &= \sup\{|\langle Tx, x \rangle| \mid x \in M^\perp, \|x\| = 1\} \\ &= \sup\{|\langle TP_{M^\perp}x, P_{M^\perp}x \rangle| \mid x \in X, \|x\| = 1\} \\ &= \sup\{|\langle P_{M^\perp}TP_{M^\perp}x, x \rangle| \mid x \in X, \|x\| = 1\} \\ &= \sup\{|\langle TP_{M^\perp}x, x \rangle| \mid x \in X, \|x\| = 1\} \\ &\quad (\text{using Lemma 3.12 and the idempotency of } P_{M^\perp}) \\ &= \|TP_{M^\perp}\| \\ &\quad (\text{since } TP_{M^\perp} \text{ is selfadjoint and using [3, Proposition 2.13, p. 34]}) \\ &= c(T), \end{aligned}$$

which proves (3.26.3). Finally, if T is also nonnegative, then $0 \leq c_1 \leq c_2$ and so $\max\{c_2, -c_1\} = c_2$. Thus (3.26.4) follows from (3.26.3). \square

Lemma 3.27. *Let T be selfadjoint and nonexpansive, and let c_1 and c_2 be defined as in (3.26.1) and (3.26.2). Then*

$$(3.27.1) \quad \|A_T(y)\| \leq \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right) \|y\| \quad \text{for every } y \in M^\perp.$$

In particular,

$$(3.27.2) \quad \|A_T^n(y)\| \leq \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right)^n \|y\| \quad \text{for every } y \in M^\perp, n \in \mathbb{N}.$$

The inequality (3.27.2) follows from (3.27.1) by induction, using the fact that $A_T(y) \in M^\perp$ whenever $y \in M^\perp$ (Lemma 3.3(4)). Our proof of (3.27.1), just like that of Theorem 3.20, uses the spectral theorem. Before proving this lemma, let us state a few consequences of it.

Theorem 3.28. *Let T be selfadjoint and nonexpansive, and let c_1 and c_2 be defined as in (3.26.1) and (3.26.2). Then*

$$(3.28.1) \quad \|A_T^{n-1}(Tx) - P_Mx\| \leq \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right)^{n-1} c(T) \|x - P_Mx\|$$

for every $x \in X$ and $n \in \mathbb{N}$.

Proof. Let $x \in X$ and set $y = Tx - P_Mx$. Then $y \in M^\perp$ by Lemma 3.3(6). Substitute this y into (3.27.2) (and replace n by $n - 1$) to obtain

$$\|A_T^{n-1}(Tx - P_Mx)\| \leq \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right)^{n-1} \|Tx - P_Mx\|.$$

But $A_T^{n-1}(Tx - P_Mx) = A_T^{n-1}(Tx) - P_Mx$ by Lemma 3.8(3) and

$$\|Tx - P_Mx\| = \|T(x - P_Mx)\| \leq c_1(T) \|x - P_Mx\| \quad \text{by (3.14.1).}$$

This proves (3.28.1). □

Theorem 3.29. *Let T be selfadjoint, nonnegative, and nonexpansive. Then*

$$(3.29.1) \quad \|A_T^{n-1}(Tx) - P_Mx\| \leq \frac{c(T)^n}{[2 - c(T)]^{n-1}} \|x - P_Mx\| \quad \text{for every } x \in X, n \in \mathbb{N}.$$

Proof. Since T is nonnegative, $c_1 \geq 0$ and $c(T) = c_2$ by Lemma 3.26. Since T is nonexpansive, $c_2 \leq 1$. Thus

$$0 \leq c_1 \leq c_2 = c(T) \leq 1.$$

Then, using Theorem 3.28, we obtain that for every $x \in X$,

$$(3.29.2) \quad \begin{aligned} \|A_T^{n-1}(Tx) - P_Mx\| &\leq \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right)^{n-1} c(T) \|x - P_Mx\| \\ &= \left(\frac{c(T) - c_1}{2 - c_1 - c(T)} \right)^{n-1} c(T) \|x - P_Mx\|. \end{aligned}$$

Now $\frac{c(T) - c_1}{2 - c_1 - c(T)} = E(c_1, c(T))$, c_1 and 0 are in the same component of the domain of $E(\cdot, c(T))$, and $E(\cdot, c(T))$ is a decreasing function by Lemma 3.25. This implies that

$$\frac{c(T) - c_1}{2 - c_1 - c(T)} = E(c_1, c(T)) \leq E(0, c(T)) = \frac{c(T)}{2 - c(T)}.$$

This together with (3.29.2) yields (3.29.1). □

Remarks. Comparing (3.29.1) with (3.16.2), we see that for each selfadjoint, nonnegative, and nonexpansive operator T , it follows that

$$(3.16.2) \quad \|A_T^{n-1}(Tx) - P_Mx\| \leq \left[\prod_{i=1}^{n-1} f(x_i) \right] c(T)^n \|x - P_Mx\|$$

and

$$(3.29.1) \quad \|A_T^{n-1}(Tx) - P_M x\| \leq \frac{c(T)^n}{[2 - c(T)]^{n-1}} \|x - P_M x\|.$$

Thus it is natural to ask whether one of these bounds is *always* better than the other. In other words, do either one of the following two inequalities *always* hold:

$$(a) \quad \prod_1^{n-1} f(x_i) \leq \frac{1}{[2 - c(T)]^{n-1}} \quad \text{for all } n \geq 2,$$

or

$$(b) \quad \frac{1}{[2 - c(T)]^{n-1}} \leq \prod_1^{n-1} f(x_i) \quad \text{for all } n \geq 2?$$

We now show that *neither* of these two inequalities always holds. To see that inequality (b) does not always hold, consider the example when $X = \ell_2(2)$ is the Euclidean plane, M_1 (resp., M_2) is the horizontal (resp., vertical) axis, and $T = P_{M_1} P_{M_2} P_{M_1}$. Then $T = 0$, $M = \text{Fix } T = \{0\}$, $c(T) = \|TP_{M^\perp}\| = 0$, $f(x) = 0$ for all $x \in \ell_2(2)$, and $\frac{1}{2-c(T)} = \frac{1}{2}$. Hence

$$\prod_1^n f(x_i) < \frac{1}{[2 - c(T)]^n} \quad \text{for every } n \geq 1.$$

To see that (a) does not always hold, let $X = \ell_2(2)$ denote the Euclidean plane and define T on X by $T(\alpha e_1 + \beta e_2) = \frac{99}{100}\alpha e_1 + \frac{19}{100}\beta e_2$. Then T is a nonnegative selfadjoint linear operator on X , $M = \text{Fix } T = \{0\}$, and $c(T) = \|T\| = \frac{99}{100}$. Letting $x_0 := \frac{10}{11}e_1 + \frac{10}{19}e_2$, we can easily deduce that $x_1 := Tx_0 = \frac{9}{10}e_1 + \frac{1}{10}e_2$, $Tx_1 = \frac{891}{1000}e_1 + \frac{19}{1000}e_2$, $t_{x_1} = \frac{\langle x_1, x_1 - Tx_1 \rangle}{\|x_1 - Tx_1\|^2} = \frac{100}{41}$, and $A_T(x_1) = t_{x_1}Tx_1 + (1 - t_{x_1})x_1 = \frac{36}{41}e_1 - \frac{4}{41}e_2$. Hence, $f(x_1) = \frac{\|A_T(x_1)\|}{\|Tx_1\|} = \frac{1000}{41}(\frac{656}{397121})^{\frac{1}{2}} = 0.9913034925\dots$ and $\frac{1}{2-c(T)} = \frac{100}{101} = 0.9900990099\dots$ implies that

$$\frac{1}{2 - c(T)} < f(x_1);$$

so (a) fails for $n = 2$.

Proof of Lemma 3.27. We should first note that $c_1 + c_2 < 2$, and hence the expressions on the right side of both (3.27.1) and (3.27.2) are well-defined. For otherwise, $c_1 = c_2 = 1$ and $\langle x, Tx \rangle = 1$ for all $x \in M^\perp$ with $\|x\| = 1$. By the condition of equality in the Schwarz inequality, this implies that $x = Tx$ for all $x \in M^\perp$. That is, $M^\perp \subset M$ and so $M^\perp = \{0\}$. But this implies that $c_1 = c_2 = 0$, a contradiction. It follows also that $E(c_1, c_2) \geq 0$.

In the notation of Lemma 3.25, we must show that

$$(3.27.2) \quad \|A_T(y)\| \leq E(c_1, c_2) \|y\| \quad \text{for every } y \in M^\perp.$$

If $M^\perp = \{0\}$, then (3.27.2) is obvious; both sides are in fact 0. Thus we can assume $M^\perp \neq \{0\}$. Fix any $y \in M^\perp \setminus \{0\}$. By scaling and Lemma 3.8(4), we may assume $\|y\| = 1$. Let

$$N := \text{span}\{y, Ty\}.$$

Then $N \subset M^\perp$ by Lemma 3.3(4) and $1 \leq \dim N \leq 2$. If $\dim N = 1$, then $Ty = \alpha y$ for some scalar $\alpha \neq 1$ and thus

$$0 \in \text{span}\{y\} = \text{span}\{y, Ty\} = \text{aff}\{y, Ty\}$$

implies $A_T(y) = 0$ since $A_T(y)$ is the point in $\text{aff}\{y, Ty\}$ having minimal norm by Theorem 3.7. Hence, (3.27.2) holds and we may therefore assume that $\dim N = 2$. In particular, $Ty \notin \text{span}\{y\}$.

Define the operator $S := P_N T P_N$. Then S is a compact selfadjoint (nonexpansive) operator with $\mathcal{R}(S) \subset N$, and thus $n := \dim \mathcal{R}(S) \leq 2$. But both y and Ty are in N ; so

$$Sy = P_N T P_N y = P_N Ty = Ty$$

implies that $Ty \in \mathcal{R}(S)$ and hence $1 \leq n \leq 2$. By the spectral theorem [3, Corollary 5.4, p. 47], there exist an orthonormal basis $\{e_i\}_1^n$ of $\mathcal{N}(S)^\perp (= \mathcal{R}(S))$ and scalars $\{\lambda_i\}_1^n$ such that

$$(3.27.3) \quad Sx = \sum_1^n \lambda_i \langle x, e_i \rangle e_i \quad \text{for every } x \in X.$$

In particular,

$$(3.27.4) \quad Se_j = \lambda_j e_j \quad (j = 1, \dots, n);$$

so each e_j is an eigenvector of S with eigenvalue λ_j . Also,

$$(3.27.5) \quad \begin{aligned} \lambda_j &= \langle \lambda_j e_j, e_j \rangle = \langle Se_j, e_j \rangle = \langle P_N T P_N e_j, e_j \rangle \\ &= \langle T P_N e_j, P_N e_j \rangle = \langle T e_j, e_j \rangle \end{aligned}$$

since each $e_j \in \mathcal{R}(S) \subset N$. Since $N \subset M^\perp$, this proves that

$$(3.27.6) \quad c_1 \leq \lambda_j \leq c_2 \quad (j = 1, \dots, n).$$

We consider two cases.

Case 1. $n = 1$.

Then since

$$N = \mathcal{R}(S) \oplus [\mathcal{R}(S)^\perp \cap N],$$

$\dim N = 2$, and $\dim \mathcal{R}(S) = 1$, it follows that $\dim[\mathcal{R}(S)^\perp \cap N] = 1$. Hence we can choose $e_2 \in \mathcal{R}(S)^\perp \cap N$ with $\|e_2\| = 1$ and define $\lambda_2 = 0$. Then $\{e_1, e_2\}$ is a basis for N , and $Se_2 = 0 = \lambda_2 e_2$. It follows that (3.27.3)–(3.27.6) hold with $n = 2$.

Case 2. $n = 2$.

Then $\mathcal{R}(S) = N$ and (3.27.3)–(3.27.6) holds with $n = 2$.

Thus each case can be reduced to the case when $n = 2$.

If $E(c_1, c_2) \geq 1$, then (3.27.2) is obvious since then

$$\|A_T(x)\| \leq \|x\| \leq E(c_1, c_2) \|x\|$$

for each x , where (3.7.5) was used for the first inequality. Thus, we may assume that $0 \leq E(c_1, c_2) < 1$. By Lemma 3.25, this is equivalent to $(1 - c_1)(1 - c_2) > 0$. That is, either $1 - c_1 > 0$ and $1 - c_2 > 0$, or $1 - c_1 < 0$ and $1 - c_2 < 0$. But the latter inequality implies $c_2 > 1$ which contradicts the nonexpansiveness of T . Thus, we must have $1 - c_1 > 0$ and $1 - c_2 > 0$. That is,

$$(3.27.7) \quad -1 \leq c_1 \leq \lambda_j \leq c_2 < 1 \quad (j = 1, 2),$$

where the lower bound $c_1 \geq -1$ is also a consequence of the nonexpansiveness of T .

Moreover, since $\{e_1, e_2\}$ is an orthonormal basis for N and since y and Ty are in N , we have $y = \sum_1^2 \alpha_i e_i$ and $Ty = Sy = \sum_1^2 \lambda_i \alpha_i e_i$, where $\alpha_i := \langle y, e_i \rangle$ ($i = 1, 2$). Then by (3.5.3) and using the fact that $\alpha_1^2 + \alpha_2^2 = \|y\|^2 = 1$, we deduce that

$$\begin{aligned} \|A_T(y)\|^2 &= \|y\|^2 - \frac{\langle y, y - Ty \rangle^2}{\|y - Ty\|^2} \\ &= 1 - \frac{\langle \sum_1^2 \alpha_i e_i, \sum_1^2 \alpha_i e_i - \sum_1^2 \lambda_i \alpha_i e_i \rangle^2}{\|\sum_1^2 \alpha_i e_i - \sum_1^2 \lambda_i \alpha_i e_i\|^2} \\ &= 1 - \frac{[\sum_1^2 \alpha_i^2 (1 - \lambda_i)]^2}{\sum_1^2 \alpha_i^2 (1 - \lambda_i)^2}. \end{aligned}$$

Putting the expression on the right over a common denominator, expanding, and simplifying, we obtain

$$(3.27.8) \quad \|A_T(y)\|^2 = \frac{\alpha_1^2 \alpha_2^2 (\lambda_2 - \lambda_1)^2}{\alpha_1^2 (1 - \lambda_1)^2 + \alpha_2^2 (1 - \lambda_2)^2}.$$

If $\lambda_1 = \lambda_2$, then (3.27.8) implies that $A_T(y) = 0$ and (3.27.2) is obvious. Thus we may assume that $\lambda_1 \neq \lambda_2$. In fact, by reindexing if necessary, we may assume that $\lambda_1 < \lambda_2$. Define $h : [0, 1] \rightarrow \mathbb{R}$ by

$$(3.27.9) \quad h(t) := \frac{t(1-t)(\lambda_2 - \lambda_1)^2}{t(1-\lambda_1)^2 + (1-t)(1-\lambda_2)^2}.$$

Since $\alpha_1^2 + \alpha_2^2 = 1$, we see that (3.27.8) implies that

$$(3.27.10) \quad \|A_T(y)\|^2 \leq \max\{h(t) \mid 0 \leq t \leq 1\}.$$

But $h(0) = h(1) = 0$ and $h(t) > 0$ for all $0 < t < 1$. Hence the maximum of h over $[0, 1]$ occurs for some $t \in (0, 1)$ that satisfies $h'(t) = 0$. Differentiating h and expanding, we deduce that

$$[ta^2 + (1-t)b^2]^2 h'(t) = (a-b)^2 [t(b-a) - b] [t(a+b) - b],$$

where $0 < b := 1 - \lambda_2 < 1 - \lambda_1 =: a$. Hence $h'(t) = 0$ if and only if $t = b/(b-a) < 0$ or $t = b/(a+b) \in (0, 1)$. Hence the maximum of h over $[0, 1]$ is attained at $t = b/(a+b)$. Thus

$$\max_{0 \leq t \leq 1} h(t) = h\left(\frac{b}{a+b}\right) = \left(\frac{a-b}{a+b}\right)^2 = \left(\frac{\lambda_2 - \lambda_1}{2 - \lambda_2 - \lambda_1}\right)^2 = E(\lambda_1, \lambda_2)^2.$$

Combining this with (3.27.10), we obtain that $\|A_T(y)\|^2 \leq E(\lambda_1, \lambda_2)^2$ or, equivalently,

$$(3.27.11) \quad \|A_T(y)\| \leq |E(\lambda_1, \lambda_2)| = E(\lambda_1, \lambda_2).$$

By Lemma 3.25, $E(\cdot, \lambda_2)$ is a decreasing function so that by (3.27.7), we get

$$(3.27.12) \quad E(\lambda_1, \lambda_2) \leq E(c_1, \lambda_2).$$

On the other hand, by Lemma 3.25, $E(c_1, \cdot)$ is an increasing function. By (3.27.7), it follows that

$$(3.27.13) \quad E(c_1, \lambda_2) \leq E(c_1, c_2).$$

Combining (3.27.11)–(3.27.13), we obtain

$$(3.27.14) \quad \|A_T(y)\| \leq E(c_1, c_2),$$

and this is just (3.27.2). □

Remarks. It is perhaps worth noting that the inequality (3.27.2), and hence the main inequality in each of Theorems 3.28 and 3.29, is *sharp*, at least for a large class of operators T . More precisely, one can prove the following result. If $T : X \rightarrow X$ is selfadjoint, nonexpansive, has finite rank, and is not the identity, then there exists $x^* \in M^\perp$ with $\|x^*\| = 1$ and

$$\|A_T^n x^*\| = \left(\frac{c_2 - c_1}{2 - c_1 - c_2} \right)^n \|x^*\| \quad \text{for } n = 0, 1, 2, \dots$$

Our proof of this result was divided into two cases: when $\mathcal{R}(T) \neq X$ and when $\mathcal{R}(T) = X$. Since the proof was somewhat lengthy, we have omitted it.

Finally, we should mention that there are examples of *expansive*, selfadjoint, and positive mappings T for which the algorithm (3.1.3) *diverges* for every nonzero x , but the accelerated counterpart (3.1.4) converges! That is, it is not always necessary to have the original algorithm converging to be able to accelerate it.

For example, let X be the Euclidean plane $\ell_2(2)$ and define $T : X \rightarrow X$ by $Tx = 3x(1)e_1 + 4x(2)e_2$. Then T is selfadjoint and positive, $M := \text{Fix } T = \{0\}$, and $\|T\| = 4$ (so T is expansive). However, $\|T^n x\| \geq 3^n \|x\|$ and $\|A_T^n(Tx)\| \leq 3^{-n+1} \|x\|$ for every x . This shows that $\|T^n x - P_M x\| \rightarrow \infty$ for each $x \neq 0$, while $\|A_T^{n-1}(Tx) - P_M x\| \rightarrow 0$ for each x .

Added in proof. Recently, there has been related work that has appeared since this paper was first submitted to the Transactions in July of 1999.

First, the authors of this paper showed that the iterates $x_0 = x$, $x_n = A_T(Tx_{n-1})$ for $n \geq 1$ generated by the accelerated map for a linear nonexpansive map T converge *weakly* to $P_{\text{Fix } T}(x)$ (*Fejér monotonicity and weak convergence of an accelerated method of projections*, Canadian Math. Soc., Conference Proceedings, **27**(2002), 1–6). This generalizes the relation (3.24.10) above.

F. Deutsch (*Accelerating the convergence of the method of alternating projections via a line search: a brief survey*, in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* (edited by D. Butnariu, Y. Censor, and S. Reich), 2001, Elsevier Science, 203–217) gave a survey of line search methods for accelerating the convergence of the method of alternating projections.

J. Xu and L. Zikatanov (*The method of alternating projections and the method of subspace corrections in Hilbert space*, J. Amer. Math. Soc., **15**(2002), 573–597) gave an identity for estimating the norm of a product of nonexpansive linear operators on a Hilbert space.

REFERENCES

1. N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68(1950), 337–404. MR **14**:479c
2. H. H. Bauschke, J. M. Borwein, and A. S. Lewis, *The method of cyclic projections for closed convex sets in Hilbert space*, Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995), 1–38, Contemporary Mathematics 204, Amer. Math. Soc., Providence, R.I., 1997. MR **98c**:49069
3. J. B. Conway, *A Course in Functional Analysis* (second edition), Graduate Texts in Mathematics 96, Springer-Verlag, New York, 1990. MR **91e**:46001

4. L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications* (second edition), Academic Press, San Diego, CA, 1999. MR **99k**:46001
5. F. Deutsch, *Rate of convergence of the method of alternating projections*, Parametric Optimization and Approximation, ISNM 72 (B. Brosowski and F. Deutsch, eds.), Birkhäuser, Basel, 1985, pp. 96-107. MR **88d**:41026
6. F. Deutsch, *The method of alternating orthogonal projections*, in Approximation Theory, Spline Functions and Applications (S. P. Singh, ed.), Kluwer Academic Publ., Dordrecht, 1992, pp. 105-121. MR **93a**:41047
7. F. Deutsch, *The angle between subspaces of a Hilbert space*, in Approximation Theory, Wavelets and Applications (S.P. Singh, ed.), Kluwer Academic Publ., Dordrecht, pp. 107-130. MR **96e**:46027
8. F. Deutsch and H. Hundal, *The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case*, Numer. Funct. Anal. and Optimiz. **15** no. 5-6 (1994), 537-565. MR **95f**:49047
9. F. Deutsch and H. Hundal, *The rate of convergence for the method of alternating projections, II*, J. Math. Anal. Appl. **205** (1997), 381-405. MR **97i**:41025
10. J. Dyer, *Acceleration of the convergence of the Kaczmarz method and iterated homogeneous transformations*, doctoral dissertation (1965).
11. C. Franchetti and W. Light, *On the von Neumann alternating algorithm in Hilbert space*, J. Math. Anal. Appl. **114** (1986), 305-314. MR **87f**:41058
12. K. Friedrichs, *On certain inequalities and characteristic value problems for analytic functions and functions of two variables*, Trans. Amer. Math. Soc. **41** (1937), 321-364.
13. W. B. Gearhart and M. Koshy, *Acceleration schemes for the method of alternating projections*, J. Comp. Appl. Math. **26** (1989), 235-249. MR **90h**:65095
14. L. G. Gubin, B. T. Polyak, and E. V. Raik, *The method of projections for finding the common point of convex sets*, USSR Computational Mathematics and Mathematical Physics **7**(6) (1967), 1-24.
15. I. Halperin, *The product of projection operators*, Acta. Sci. Math. (Szeged) **23** (1962), 96-99. MR **25**:5373
16. M. Hanke and W. Niethammer, *On the acceleration of Kaczmarz's method for inconsistent linear systems*, Linear Algebra Appl. **130** (1990), 83-98. MR **91f**:65065
17. S. Kayalar and H. Weinert, *Error bounds for the method of alternating projections*, Math. Control Signals Systems **1** (1988), 43-59. MR **89b**:65137
18. J. von Neumann, *Functional Operators. II*, Princeton University Press, Princeton, NJ, 1950. [This is a reprint of mimeographed lecture notes first distributed in 1933.] MR **11**:599e
19. F. Riesz and B. Sz.-Nagy, *Über Kontraktionen des Hilbertschen Raumes*, Acta. Sci. Math. **10** (1941-1943), 202-205. MR **8**:35a
20. F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Ungar, New York, 1955. MR **17**:175i
21. R. Smarzewski, *Iterative recovering of orthogonal projections*, preprint (December, 1996).
22. K. T. Smith, D. C. Solmon, and S. L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. **83** (1976), 1227-1270. MR **58**:9394a

DEPARTMENT OF MATHEMATICS AND STATISTICS, OKANAGAN UNIVERSITY COLLEGE, KELOWNA, BRITISH COLUMBIA, CANADA V1V 1V7

E-mail address: bauschke@cecm.sfu.ca

Current address: Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

DEPARTMENT OF MATHEMATICS, THE PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PENNSYLVANIA 16802

E-mail address: deutsch@math.psu.edu

NONRAND, 12100 WILTSHIRE #1650, LOS ANGELES, CALIFORNIA 90025

E-mail address: hundalhm@vicon.net

Current address: 146 Cedar Ridge Drive, Port Matilda, Pennsylvania 16870

DEPARTMENT OF MATHEMATICS, SOGANG UNIVERSITY, SEOUL, KOREA

E-mail address: shpark@ccs.sogang.ac.kr

ANDERSON'S DOUBLE COMPLEX AND GAMMA MONOMIALS FOR RATIONAL FUNCTION FIELDS

SUNGHAN BAE, ERNST-ULRICH GEKELER, PYUNG-LYUN KANG,
 AND LINSHENG YIN

ABSTRACT. We investigate algebraic Γ -monomials of Thakur's positive characteristic Γ -function, by using Anderson and Das' double complex method of computing the sign cohomology of the universal ordinary distribution. We prove that the Γ -monomial associated to an element of the second sign cohomology of the universal ordinary distribution of $\mathbb{F}_q(T)$ generates a Kummer extension of some Carlitz cyclotomic function field, which is also a Galois extension of the base field $\mathbb{F}_q(T)$. These results are characteristic- p analogues of those of Deligne on classical Γ -monomials, proofs of which were given by Das using the double complex method. In this paper, we also obtain some results on e -monomials of Carlitz's exponential function.

0. INTRODUCTION

In [An1] Anderson invented a remarkable method of computing in an identical way the sign cohomology of the universal ordinary distributions, both for the rational number field and a global function field. He introduced a certain double complex which is a resolution of the universal ordinary distribution. This double complex enabled him to construct canonical basis classes of the sign cohomology. Das [Da] used this double complex in the rational number field case for the study of classical Γ -monomials and got a series of results, which greatly illuminated the power of Anderson's method.

In this paper, using Anderson's double complex and following Das' method, we study Γ -monomials for rational function fields. Thakur [Th] defined the Γ -function in characteristic p and showed that it has many interesting properties analogous to the classical Γ -function. Especially, it satisfies a reflection formula and a multiplication formula. Sinha [Si] used Anderson's soliton theory to develop an analogue of Deligne's reciprocity for function fields. In the course of this he found that certain Γ -monomials generate Kummer extensions of cyclotomic function fields, a result which will be reproved below with the aid of the double complex. Using Γ -monomials we also find extensions of cyclotomic function fields, and these happen to be Galois even over the basic rational function field.

We would like to emphasize the following technical points: Besides the double complex, there are several main ingredients in computing the Γ -monomials in Das'

Received by the editors March 12, 2001.

2000 *Mathematics Subject Classification.* Primary 11R58.

The first author was supported by KOSEF cooperative Research Fund and DFG.

The fourth author was supported by Distinguished Young Grant in China and a fund from Tsinghua.

paper, and these are used frequently. In the case of a rational function field there are more roots of unity, which causes the definitions of the vertical shift operator and “canonical lifting operator” to be more complicated. In addition, the reflection formula and the multiplication formula of the Γ -function play important roles in our study. These formulae in the function field case have some extra factors, and thus one has to be more careful in applying them.

1. THE DOUBLE COMPLEX FOR $\mathbb{F}_q(T)$

Let $K = \mathbb{F}_q(T)$ and $A = \mathbb{F}_q[T]$, the rational function field and polynomial ring, respectively, over the finite field \mathbb{F}_q . We fix a generator γ of $J = \mathbb{F}_q^*$. Let \mathcal{A} be the free abelian group generated by symbols $[a]$ with $a \in K/A$. Let \mathbb{U} be the quotient of \mathcal{A} by the subgroup generated by all elements $[a] - \sum_{\mathbf{n}b=a} [b]$, where \mathbf{n} is a monic polynomial in A , and \mathbb{U}^- (resp. \mathbb{U}^+) the quotient of \mathcal{A} by the subgroup generated by all elements $[a] - \sum_{\mathbf{n}b=a} [b]$, along with all the $\sum_{\theta \in J} [\theta a]$ (resp. $[a] - [\gamma a]$). We call the group \mathbb{U} the universal ordinary distribution on K/A . Further, J acts on \mathbb{U} in the natural way. Let $H^*(J, \mathbb{U})$ denote the sign cohomology group for \mathbb{U} . It is known that $\text{tor}(\mathbb{U}^+) \simeq H^1(J, \mathbb{U})$ and $\text{tor}(\mathbb{U}^-) \simeq H^2(J, \mathbb{U})$ ([BGY], Proposition 2.4). If $\mathbf{a} = \sum m_i [a_i] \in \mathcal{A}$ represents an element in $H^*(J, \mathbb{U})$, we often write $\mathbf{a} \in H^*(J, \mathbb{U})$. It is clear from the context whether elements of \mathcal{A} , \mathbb{U} , $H^1(J, \mathbb{U})$, or $H^2(J, \mathbb{U})$ are intended. We use gothic letters to denote elements of A . Define

$$\left\langle \frac{\mathbf{a}}{\mathbf{f}} \right\rangle = \begin{cases} 1, & \text{if } \mathbf{a} \text{ is monic} \\ 0, & \text{otherwise,} \end{cases}$$

assuming that $\deg \mathbf{a} < \deg \mathbf{f}$ and that \mathbf{f} is monic. For $\mathbf{a} = \sum m_i [a_i] \in \mathcal{A}$ we define the *total sum* $TS(\mathbf{a})$ and *internal sum* $IS(\mathbf{a})$ of \mathbf{a} by $\sum m_i$ and by $IS(\mathbf{a}) = \sum m_i \langle a_i \rangle$, respectively. Let \mathbf{f} be the least common multiple of the denominators of the a_i and let $\mathbf{t} \in (A/\mathbf{f})^*$. We define $\mathbf{a}^{\mathbf{t}}$ by

$$\mathbf{a}^{\mathbf{t}} = \sum m_i [\mathbf{t}a_i].$$

Let \mathcal{P} be the set of all monic irreducible polynomials in A . We fix a linear order “ $<$ ” on \mathcal{P} . Let

$$\mathcal{S} = \{[a, \mathbf{g}, n] : a \in K/A, \mathbf{g} \text{ a squarefree monic polynomial, } n \text{ an integer}\}.$$

We denote by $|\mathbf{g}|$ the number of monic irreducible polynomials dividing \mathbf{g} . We define a double complex \mathbb{SK} as follows: $\mathbb{SK}_{m,n}$ = the free abelian group generated by the symbols $[a, \mathbf{g}, n] \in \mathcal{S}$ with $m = |\mathbf{g}|$. The chain maps ∂ and δ of bidegree $(-1, 0)$ and $(0, -1)$, respectively, are defined by

$$\partial[a, \mathbf{g}, n] = \sum_{i=1}^{|\mathbf{g}|} (-1)^{i-1} ([a, \mathbf{g}/\mathbf{p}_i, n] - \sum_{\mathbf{p}_i b=a} [b, \mathbf{g}/\mathbf{p}_i, n]),$$

where $\mathbf{g} = \mathbf{p}_1 \cdots \mathbf{p}_m$ with $\mathbf{p}_i < \mathbf{p}_j$ for $i < j$, and

$$\delta[a, \mathbf{g}, n] = \begin{cases} (-1)^m \sum_{i=0}^{q-2} [\gamma^i a, \mathbf{g}, n-1], & \text{for } n \text{ odd,} \\ (-1)^m ([a, \mathbf{g}, n-1] - [\gamma a, \mathbf{g}, n-1]), & \text{for } n \text{ even.} \end{cases}$$

Then it is easy to see that

$$\partial^2 = 0, \quad \delta^2 = 0 \quad \text{and} \quad \delta\partial + \partial\delta = 0.$$

Let $(T(\mathbb{SK}), \partial + \delta)$ be the total complex of \mathbb{SK} . We use the same notation \mathbb{SK} for the total complex when the meaning is evident.

Let \mathbb{SK}' be the subcomplex of \mathbb{SK} generated by the elements $\beta(a, n)[a, \mathfrak{g}, n]$, where

$$\beta(a, n) = \begin{cases} q-1, & \text{if } a = 0 \text{ and } n \text{ is even,} \\ 1, & \text{otherwise.} \end{cases}$$

Then following the method employed by Ouyang in [Ou], we have:

Proposition 1. *Let \mathbb{U} be the universal ordinary distribution on K/A . There exist canonical isomorphisms*

$$H^2(J, \mathbb{U}) = H_0(H_0(\mathbb{SK}, \partial), \delta) = H_0(\mathbb{SK}, \partial + \delta) = H_0(\mathbb{SK}/\mathbb{SK}', \partial + \delta)$$

and

$$H^1(J, \mathbb{U}) = H_{-1}(H_0(\mathbb{SK}, \partial), \delta) = H_{-1}(\mathbb{SK}, \partial + \delta) = H_{-1}(\mathbb{SK}/\mathbb{SK}', \partial + \delta).$$

Since

$$\delta([0, \mathfrak{g}, n]) = \begin{cases} (-1)^n(q-1)[0, \mathfrak{g}, n-1], & \text{if } n \text{ is odd,} \\ 0, & \text{if } n \text{ is even,} \end{cases}$$

and $\partial([0, \mathfrak{g}, n])$ lies in \mathbb{SK}' , we have:

Proposition 2. *Given a square-free monic polynomial \mathfrak{g} with $|\mathfrak{g}| = i$, we define*

$$k_{\mathfrak{g}} = \begin{cases} [0, \mathfrak{g}, -i] \in \mathbb{SK}_{i,-i}/\mathbb{SK}'_{i,-i}, & \text{for } i \text{ even,} \\ [0, \mathfrak{g}, -i-1] \in \mathbb{SK}_{i,-i-1}/\mathbb{SK}'_{i,-i-1}, & \text{for } i \text{ odd.} \end{cases}$$

Then the collection $\{k_{\mathfrak{g}}; |\mathfrak{g}| \text{ even}\}$ (resp. $\{k_{\mathfrak{g}}; |\mathfrak{g}| \text{ odd}\}$) forms a $\mathbb{Z}/(q-1)$ -basis for $H_0(\mathbb{SK}/\mathbb{SK}', \partial + \delta)$ (resp. $H_{-1}(\mathbb{SK}/\mathbb{SK}', \partial + \delta)$). The $k_{\mathfrak{g}}$ are referred to as canonical basis classes.

Define the vertical shift operator $S : \mathbb{SK}_{m,n} \longrightarrow \mathbb{SK}_{m,n+1}$ by the rule

$$S([a, \mathfrak{g}, n]) = (-1)^{|\mathfrak{g}|} \begin{cases} [a, \mathfrak{g}, n+1], & \text{if } n \text{ is even,} \\ -\sum_{i=1}^{q-2} i[\gamma^i a, \mathfrak{g}, n+1], & \text{if } n \text{ is odd,} \end{cases}$$

and define the diagonal shift operator $\Delta_{\mathfrak{p}} : \mathbb{SK}_{m,n} \longrightarrow \mathbb{SK}_{m-1,n+2}$ associated with a prime \mathfrak{p} by the rule:

$$\Delta_{\mathfrak{p}}([a, \mathfrak{g}, n]) = 0, \quad \text{if } \mathfrak{p} \nmid \mathfrak{g},$$

and if $\mathfrak{g} = \mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_m$ with $\mathfrak{p}_1 < \mathfrak{p}_2 < \cdots < \mathfrak{p}_m$,

$$\Delta_{\mathfrak{p}_r}([a, \mathfrak{g}, n]) = (-1)^r [a, \mathfrak{g}/\mathfrak{p}_r, n+2].$$

The reader can check directly the following lemma, or refer to [Da, Thms. 4-5].

Lemma 3. i) $S\delta + \delta S = q-1$ and $\partial S + S\partial = 0$. Thus $(\partial + \delta)S + S(\partial + \delta) = q-1$.
ii) $\partial \Delta_{\mathfrak{p}} = \Delta_{\mathfrak{p}} \partial$ and $\delta \Delta_{\mathfrak{p}} = \Delta_{\mathfrak{p}} \delta$.

Given a canonical basis class $[0, \mathfrak{g}, -n]$ with $|\mathfrak{g}| = n$ even, one can construct a representing cycle

$$C = \bigoplus_{i=0}^n C_{i,-i}, \quad C_{n,-n} = [0, \mathfrak{g}, -n],$$

such that $C_{i,-i} = \sum n_j [a_j, \mathfrak{g}_j, -i]$ with $\text{sgn}(a_j) = 1$ for i odd, $\text{sgn}(a_j) \neq \gamma^{q-2}$ for i even, and no term of the form $[0, \mathfrak{h}, -m]$ except $[0, \mathfrak{g}, -n]$ occurs, as follows.

Suppose that one has constructed $C_{i,-i}$. If $\partial C_{i,-i} = \sum_j m_j [a_j, \mathfrak{g}_j, -i]$, then

$$C_{i-1,1-i} = (-1)^{i-1} \begin{cases} \sum_j m_j \langle a_j \rangle [a_j, \mathfrak{g}, 1-i], & \text{if } i \text{ is even,} \\ \sum_j m_j \sum_{k \geq 0}^{\kappa(a_j)-1} [\gamma^{k-\kappa(a_j)} a_j, \mathfrak{g}, 1-i], & \text{if } i \text{ is odd,} \end{cases}$$

where $\text{sgn}(a_j) = \gamma^{\kappa(a_j)}$ with $0 \leq \kappa(a_j) < q-1$.

We call C a *semi-canonical lifting*. Such a construction also works for canonical basis classes of H^1 and for the boundary elements of \mathbb{SK} .

For an element C of \mathbb{SK} and a square-free monic polynomial \mathfrak{g} , we let $C^{\{\mathfrak{g}\}}$ be the \mathfrak{g} -component, i.e., the part that includes those of the form $[\ast, \mathfrak{g}, \ast]$. Following the same lines as Proposition 7 of [Da], we have:

Proposition 4. *Let $C = \bigoplus_{i+j=\ell} C_{i,j}$ be a cycle in \mathbb{SK} . For a fixed monic square-free polynomial \mathfrak{g} , write $C_{k,\ell-k}^{\{\mathfrak{g}\}} = \sum n_i [a_i, \mathfrak{g}, \ell-k]$. Then we have*

$$\sum n_i [a_i] \in \begin{cases} H^1(J, \mathbb{U}) & \text{if } \ell-k \text{ is odd,} \\ H^2(J, \mathbb{U}) & \text{if } \ell-k \text{ is even.} \end{cases}$$

The assertions in the next proposition are the analogues of Theorem 8 and Propositions 3 and 4 in [Da]. The ideas of the proof are taken from there.

Proposition 5. *Let $\mathbf{a} = \sum m_i [a_i] \in H^2(J, \mathbb{U})$. We have*

- i) *For each $\mathfrak{t} \in (A/\mathfrak{f})^*$, $\sum m_i \langle a_i \rangle = \sum m_i \langle \mathfrak{t} a_i \rangle$.*
- ii) *Let m be the coefficient of $[0]$ in \mathbf{a} . Then $q-1$ divides $\sum m_i - m$.*
- iii) *Let C be a cycle in \mathbb{SK} such that $C = \bigoplus_{i+j=0} C_{i,j}$, $C_{0,0} = \sum m_i [a_i, 1, 0]$, and $(\partial + \delta)C = 0$. Let further*

$$C_{1,-1} = \sum n_j [b_j, \mathfrak{p}_j, -1].$$

Then $\sum n_j \deg \mathfrak{p}_j \equiv 0 \pmod{q-1}$.

2. ALGEBRAIC GAMMA MONOMIALS

Thakur ([Th]) defined some Γ -function in characteristic p . We change Thakur’s definition slightly by the formula

$$\Gamma(z) = \prod_{\mathfrak{a} \in A_+} \left(1 + \frac{z}{\mathfrak{a}}\right)^{-1},$$

where A_+ is the set of all monic polynomials in A . This $\Gamma(z)$ is just the $\Pi(z)$ of Thakur. Let $\tilde{\pi}$ denote the fundamental period of the Carlitz module, which is unique up to a factor of \mathbb{F}_q^* . Let $e = e_C$ be the Carlitz exponential. The Γ -function has the following nice properties.

Theorem 6 ([Th], Theorem 6.1.1, Theorem 6.2.1). (1) Reflection formula:

$$\prod_{\theta \in J} \Gamma(\theta z) = \frac{\tilde{\pi} z}{e(\tilde{\pi} z)}.$$

(2) Multiplication formula: *For $\mathfrak{f} \in A_+$ of degree d we have*

$$\prod_{\substack{\mathfrak{a} \in A \\ \deg \mathfrak{a} < d}} \Gamma\left(\frac{z + \mathfrak{a}}{\mathfrak{f}}\right) = \tilde{\pi}^{(q^d-1)/(q-1)} ((-1)^d \mathfrak{f})^{q^d/(1-q)} R_d(z) \Gamma(z),$$

where $R_d(z) = \prod_{\substack{\deg \mathfrak{a} < d \\ \mathfrak{a} \text{ monic}}} (z + \mathfrak{a})$.

For $a \in K/A$ we denote by $\{a\}$ the representative of a such that $|a|_\infty < 1$, where $| \cdot |_ \infty$ is the absolute value at $\infty = (\frac{1}{T})$. For each $\mathbf{a} = \sum m_i[a_i] \in \mathcal{A}$, we define the Γ -monomial, e -monomial, and r -monomial, respectively, by

$$\Gamma(\mathbf{a}) = \tilde{\pi}^{\frac{TS(\mathbf{a})}{q-1}} \prod \Gamma(\{a_i\})^{-m_i},$$

$$e(\mathbf{a}) = \prod_{a_i \neq 0} e(\tilde{\pi}a_i)^{m_i}, \quad \text{and} \quad r(\mathbf{a}) = \prod_{a_i \neq 0} \{a_i\}^{m_i}.$$

By abuse of notation, we also write $\Gamma(\sum n_i[a_i, *, *])$ to mean $\Gamma(\sum n_i[a_i])$. This notation will also be applied to e - and r -monomials. In what follows $a \in K/A$ always means that $a = \{a\}$ unless otherwise stated. It is known that $\Gamma(\mathbf{a})$ is algebraic over K if $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$. In fact, we have shown that $\Gamma(\mathbf{a})^{q-1} = {}^q\sqrt{r}e(\mathbf{a})$ for some $r \in K^*$; see [BGY, Thm. 7.2]. Using the double complex, we can express r explicitly. Let $C = \bigoplus_{i+j=0} C_{i,j}$ be a cycle in $\mathbb{S}\mathbb{K}$ such that $C_{0,0} = \sum m_i[a_i, 1, 0]$ and $(\partial + \delta)C = 0$. Then $(\partial + \delta)SC = (q-1)C$ and $(q-1)C_{0,0} = \delta SC_{0,0} + \partial SC_{1,-1}$. Note that $\Gamma((q-1)C_{0,0}) = \Gamma(\mathbf{a})^{q-1}$ and $\Gamma(\delta SC_{0,0}) = e(\mathbf{a})/r(\mathbf{a})$. We get

$$(2.1) \quad \Gamma(\mathbf{a})^{q-1} = \frac{\Gamma(\partial SC_{1,-1})}{r(\mathbf{a})} e(\mathbf{a}).$$

Now the following Kummer property of Γ -monomials, which originally is due to Sinha([Si]), is a direct result of the equality. We denote by $K_{\mathfrak{f}}$ the cyclotomic function field of conductor \mathfrak{f} .

Theorem 7. *Let $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$ and let \mathfrak{f} be the least common multiple of the denominators of \mathbf{a} . Then $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))/K_{\mathfrak{f}}$ is a Kummer extension.*

Proof. For any irreducible polynomial \mathfrak{p} with degree d , we have

$$\partial S[b, \mathfrak{p}, -1] = \sum_{i=1}^{q-2} i([\gamma^i b, 1, 0] - \sum_{\deg u < d} [\frac{\gamma^i b + u}{\mathfrak{p}}, 1, 0])$$

and thus

$$\Gamma(\partial S[b, \mathfrak{p}, -1]) = ((-1)^d \mathfrak{p})^{\frac{q^d(q-2)}{2}} \cdot \left(\prod_{i=1}^{q-2} R_d(\gamma^i \{b\})^{-i} \right).$$

If $2 \mid q$, then $\Gamma(\partial S[b, \mathfrak{p}, -1]) \in K$. If $2 \nmid q$, since ${}^{q-1}\sqrt{(-1)^d \mathfrak{p}} \in K_{\mathfrak{p}}$, we get the result by Eq (2.1). \square

The key point in the study of Γ -monomials by means of the double complex is that the factor $\Gamma(\partial SC_{1,-1})$ in (2.1) is very simple, if C is a canonically lifted cycle, and for a general cycle C we get information about that factor using homological algebra. In fact, we have the following theorem and corollary, the lines of proof of which are again taken from [Da, Sect. 9].

Theorem 8. *Let n be an even positive integer. Let $C = \bigoplus C_{i,-i}$ be the semi-canonically lifted cycle from the basis class $[0, \mathfrak{g}, -n]$, where \mathfrak{g} is a square-free monic polynomial divisible by n irreducible polynomials. Let $\mathbf{a} = \sum m_i[a_i]$, where $C_{0,0} = \sum m_i[a_i, 1, 0]$. Then $\Gamma(\mathbf{a})^{q-1} \in K_{\mathfrak{g}}$. Furthermore,*

i) *If $\mathfrak{g} = \mathfrak{p}\mathfrak{q}$ with $d = \deg \mathfrak{p}$ and $e = \deg \mathfrak{q}$, then*

$$\Gamma(\mathbf{a})^{q-1} \equiv \sqrt{\frac{\mathfrak{q}^d}{\mathfrak{p}^e}} e(\mathbf{a}) \pmod{K^*}.$$

ii) If $n \geq 4$, then

$$\Gamma(\partial SC_{1,-1}) \in K^* \quad \text{and} \quad \Gamma(\mathbf{a})^{q-1} \equiv e(\mathbf{a}) \pmod{K^*}.$$

Corollary. Let $n \geq 4$ be an even integer. Let $\mathbf{a} = \sum m_i[a_i]$ represent the basis class $[0, \mathfrak{g}, -n]$ with $|\mathfrak{g}| = n$, not necessarily a semi-canonical representative. Then

$$\Gamma(\mathbf{a})^{q-1} \equiv e(\mathbf{a}) \pmod{K^*}.$$

The following result is an analogue of Theorem 11 in [Da].

Proposition 9. Let $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$. Then, with notation as in Proposition 5, we have

$$\frac{e(\mathbf{a})}{e(\mathbf{a}^t)} = \theta_t e(\mathbf{b})^{(q-1)},$$

for some $\mathbf{b} \in \mathcal{A}$ and $\theta_t = \pm 1$.

Proof. Let $C = \bigoplus_{i+j=0} C_{i,j}$ be a cycle in $\mathbb{S}\mathbb{K}$ such that $C_{0,0} = \sum m_i[a_i, 1, 0]$. Then $C - C^t$ is a boundary. Let $B = \bigoplus_{i+j=1} B_{i,j}$ be a chain in $\mathbb{S}\mathbb{K}$ such that $(\partial + \delta)B = C - C^t$. We have $\frac{e(\mathbf{a})}{e(\mathbf{a}^t)} = e(\partial B_{1,0})e(\delta B_{0,1})$. Note that

$$e(\partial[0, \mathfrak{p}, 0]) = \mathfrak{p} = \prod_{\substack{\deg \mathfrak{u} < \deg \mathfrak{p} \\ \text{monic}}} (-1)^{\frac{q \deg \mathfrak{p} - 1}{q-1}} e\left(\frac{\mathfrak{u}}{\mathfrak{p}}\right)^{q-1},$$

and

$$e(\partial[a, \mathfrak{p}, 0]) = 1, \quad \text{if } a \neq 0,$$

and that $e(\delta[a, 1, 1]) = -e(a)^{q-1}$. We get the result. \square

Remark. As the referee points out, if one defines

$$\sin a = {}^{q-1}\sqrt{-1} \cdot e(\{a\}/\text{sgn}\{a\})$$

for $a \in K \setminus \mathbb{A}$, then, by making the obvious definition of $\sin \mathbf{a}$, one has

$$\frac{\sin \mathbf{a}}{\sin \mathbf{a}^t} = (\sin \mathbf{b})^{q-1},$$

in strict analogy with Theorem 11 of [Da].

Example. Let $q = 3$, $\mathbf{a} = [\frac{1}{T+1}] - [\frac{T-1}{T(T+1)}]$, and $\mathbf{t} = -T + 1$. Then

$$\frac{e(\mathbf{a})}{e(\mathbf{a}^t)} = \frac{e(\tilde{\pi} \frac{1}{T(T+1)})}{e(\tilde{\pi} \frac{T-1}{T(T+1)})} = -e(\tilde{\pi} \frac{1}{T(T+1)})^2,$$

since $\lambda = e(\tilde{\pi} \frac{1}{T(T+1)})$ satisfies the relation $\lambda^4 + (T+1)\lambda^2 + 1 = 0$.

If $\mathbf{t} = -1$, then

$$\frac{e(\mathbf{a})}{e(\mathbf{a}^t)} = 1 = e(\mathbf{0})^2.$$

Thus θ_t changes as \mathbf{t} varies.

Theorem 10. Let $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$. Let \mathfrak{f} be the least common multiple of the denominators of the a_i and let $\mathbf{t} \in (A/\mathfrak{f})^*$. Then

$$\frac{\Gamma(\mathbf{a})}{\Gamma(\mathbf{a}^t)} \in K_{\mathfrak{f}}.$$

Proof. With notation as in the proof of Proposition 9, let $B_{1,0} = \sum \ell_j [c_j, \mathfrak{p}_j, 0]$. We may assume that B is a semi-canonically lifted chain. Then the denominators of c_j divide \mathfrak{f} . From the proof of Proposition 9, it suffices to show that $\Gamma(\partial B_{1,0}) \in K_{\mathfrak{f}}$. It can be easily checked that

$$\Gamma(\partial B_{1,0}) \equiv (-1)^{\frac{\sum \ell_j \deg \mathfrak{p}_j}{q-1}} \prod \mathfrak{p}_j^{\frac{\ell_j}{q-1}} \pmod{K_{\mathfrak{f}}^*}.$$

Thus the result follows. \square

Similarly, if \mathbf{a} and \mathbf{a}' represent the same class in $H^2(J, \mathbb{U})$, then $\frac{\Gamma(\mathbf{a})}{\Gamma(\mathbf{a}')} \in K_{\mathfrak{f}}$.

3. CRITERION FOR $H^1(J, \mathbb{U})$

The following conclusion is shown in [BGY, Cor. 4.2].

Lemma 11. *Let $\mathbf{b} = \sum m_i [b_i] \in H^1(J, \mathbb{U})$. Then for all i , $b_i \neq 0$.*

Lemma 12. *Let $\mathbf{b} = \sum m_i [b_i] \in H^1(J, \mathbb{U})$, representing any canonical basis class of $H^1(J, \mathbb{U})$ indexed by a monic square-free polynomial divisible by at least three primes. Let $C = \bigoplus_{i+j=1} C_{i,j}$ be a cycle such that $C_{0,1} = \sum m_i [b_i, 1, 1]$. Assume that no term of the form $[0, \mathfrak{p}, 0]$ appears in $C_{1,0}$. Then*

$$\sum m_i \equiv 0 \pmod{(q-1)}.$$

Proof. Note that $IS(\partial[b, \mathfrak{p}, 0]) = \frac{q^{\deg \mathfrak{p}} - 1}{q-1} \equiv \deg \mathfrak{p} \pmod{(q-1)}$ for $b \notin A$. Now follow the proof of Proposition 13 of [Da]. \square

It is shown in [BGY] that $\mathbf{b} \in H^1(J, \mathbb{U})$ if and only if $|r(\mathbf{b})|_{\infty} = |r(\mathbf{b}^{\mathfrak{t}})|_{\infty}$ for any $\mathfrak{t} \in (A/\mathfrak{f})^*$, where \mathfrak{f} is the least common multiple of the denominators of the b_i . Here we give another proof of the necessity of this using the double complex. In this way one can get some more information about the e -monomials.

Theorem 13. *Let $\mathbf{b} = \sum m_i [b_i] \in H^1(J, \mathbb{U})$ and let \mathfrak{f} be the least common multiple of the denominators of the b_i . Then*

$$|r(\mathbf{b})|_{\infty} = |r(\mathbf{b}^{\mathfrak{t}})|_{\infty},$$

for all $\mathfrak{t} \in (A/\mathfrak{f})^*$. Furthermore, we have more information about e -monomials in the following special cases:

First case: If \mathbf{b} represents a canonical basis class of $H^1(J, \mathbb{U})$, indexed by a single irreducible polynomial \mathfrak{p} , then $e(\mathbf{b})^{(q-1)} = e(\mathbf{b}^{\mathfrak{t}})^{(q-1)} = \sqrt{\pm \mathfrak{p}}$.

Second case: Let \mathbf{b} represent a canonical basis class of $H^1(J, \mathbb{U})$ indexed by a monic square-free polynomial divisible by at least three primes. Let $C = \bigoplus C_{i,-i+1}$ be a cycle such that $C_{0,1} = \sum m_i [b_i, 1, 1]$. Assume that no term of the form $[0, \mathfrak{p}, 0]$ appears in $C_{1,0}$. Then $e(\mathbf{b}^{\mathfrak{t}}) \in \mathbb{F}_q^*$ for any $\mathfrak{t} \in (A/\mathfrak{f})^*$.

Proof. The first statement follows by linearity from the two special cases, since $|r(\mathbf{b})|_{\infty} = |e(\mathbf{b})|_{\infty}$. Let $C = \bigoplus C_{i,-i+1}$ be a cycle such that $C_{0,1} = \sum m_i [b_i, 1, 1]$.

First Case: Let $C_{1,0} = [0, \mathfrak{p}, 0]$. Then we know that $e(-\partial C_{1,0}) = e(-\partial C_{1,0}^{\mathfrak{t}}) = \mathfrak{p}$. Thus

$$e(\delta C_{0,1}) = e(\delta C_{0,1}^{\mathfrak{t}}) = \mathfrak{p}.$$

However, $e(\delta C_{0,1}) = (-1)^{\sum m_i} e(\mathbf{b})^{q-1}$ and $e(\delta C_{0,1}^{\mathfrak{t}}) = (-1)^{\sum m_i} e(\mathbf{b}^{\mathfrak{t}})^{q-1}$. Then the result follows from the fact that $\sigma_{\mathfrak{t}} e(\mathbf{b}) = e(\mathbf{b}^{\mathfrak{t}})$.

Second Case: In this case, $e(\partial C_{1,0}) = 1$. Then the result follows in the same way as the first case, by using Lemma 11. \square

The following proposition is a direct consequence of Theorem 13.

Proposition 14. *Let $\mathbf{b} = \sum m_i[b_i] \in H^1(J, \mathbb{U})$. Assume that \mathbf{b} represents a canonical basis class of $H^1(J, \mathbb{U})$, indexed by monic square-free polynomials divisible by at least three monic irreducibles. Let C be a cycle in \mathbb{SK} such that $C = \bigoplus_{i+j=0} C_{i,j}$, $C_{0,0} = \sum m_i[b_i, 1, 0]$ and $(\partial + \delta)C = 0$. Assume that no term of the form $[0, \mathfrak{p}, -1]$ appears in $C_{1,-1}$. Then for each \mathfrak{t} coprime to the least common multiple of the denominators of the b_i , we have*

$$\prod_i \operatorname{sgn}(\{b_i \mathfrak{t}\})^{m_i} = \prod_i \operatorname{sgn}(\{b_i\})^{m_i}.$$

Proof. From Theorem 13, we know that $e(\mathbf{b}) \in K^*$. Thus $e(\mathbf{b}^{\mathfrak{t}}) = \sigma_{\mathfrak{t}}e(\mathbf{b}) = e(\mathbf{b})$ and, further, $\operatorname{sgn}(e(\mathbf{b})) = \operatorname{sgn}(e(\mathbf{b}^{\mathfrak{t}}))$. Since $\sum m_i \equiv 0 \pmod{q-1}$ from Lemma 12, we get the result. \square

4. GALOIS PROPERTIES OF $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))/K$

Let $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$, and let \mathfrak{f} be the least common multiple of the denominators of the a_i . In this section we consider the extension $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))$ over K . Let $C = \bigoplus C_{i,-i}$ be a cycle in \mathbb{SK} such that $C_{0,0} = \sum m_i[a_i, 1, 0]$. Write

$$v = \Gamma(\partial SC_{1,-1}) \quad \text{and} \quad \Gamma(\mathbf{a})^{q-1} = ve(\mathbf{a})/r(\mathbf{a}).$$

Let σ be an element of $\operatorname{Gal}(\bar{K}/K)$ whose restriction to $K_{\mathfrak{f}}$ is $\sigma_{\mathfrak{t}}$, where \bar{K} is the separable closure of K . Then

$$\left(\frac{\Gamma(\mathbf{a})}{\sigma\Gamma(\mathbf{a})}\right)^{q-1} = \frac{v}{\sigma v}\theta_{\mathfrak{t}}e(\mathbf{b})^{q-1},$$

where $\theta_{\mathfrak{t}}$ and \mathbf{b} are given in Proposition 9. Hence

$$\frac{\Gamma(\mathbf{a})}{\sigma\Gamma(\mathbf{a})} \in K_{\mathfrak{f}} \iff \frac{v}{\sigma v}\theta_{\mathfrak{t}} = 1.$$

When is $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))/K$ a Galois extension? The following theorems are the main results of the paper.

Theorem 15. *Assume q is odd. Let $\mathbf{a} = \sum m_i[a_i] \in H^2(J, \mathbb{U})$, and let \mathfrak{f} be the least common multiple of the denominators of the a_i . Then $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))$ is a Galois extension of K .*

In the number field case, Das first shows the analogue of Theorem 8' [Da, Thm. 15(C)] and then derives easily his main result [Da, Thm. 16]. In our case, Theorem 8' is hard to prove because of the extra factor in the functional equation of the Γ -function. So we first show this theorem and then derive Theorem 8'. In the following proof we omit some complicated computations as suggested by the referee, because these are not interesting and do not offer any deeper insights.

Proof. The notation is as above. Using the analogue of [Da, Prop. 16] we can assume that the cycle C that \mathbf{a} represents is semi-canonically lifted from a basis class. We know from Proposition 5 that $e(\mathbf{a})$ and $e(\mathbf{a}^{\mathfrak{t}})$ lie in $K_{\mathfrak{f}}^+$, the maximal real subfield of $K_{\mathfrak{f}}$. Hence the signs of $e(\mathbf{a})$ and $e(\mathbf{a}^{\mathfrak{t}})$ make sense. It is not hard to get $\operatorname{sgn}(e(\mathbf{a})/e(\mathbf{a}^{\mathfrak{t}})) = 1$.

Let B be a semi-canonically lifted chain such that $C - C^t = (\partial + \delta)B$. Let $B_{0,1} = \sum n_j [b_j, 1, 1]$ and let $\mathbf{b} = \sum n_j [b_j]$. We see that $\text{sgn}(e(\mathbf{b})^{q-1}) = (-1)^{\sum n_j}$ and thus $\theta_t = (-1)^{\sum n_j}$ by Proposition 9.

Write $B_{1,0} = \sum l_k [c_k, \mathbf{p}_k, 0]$. Using the facts that $\partial B_{1,0} + \delta B_{0,1} = C_{0,0} - C_{0,0}^t$ and $IS(C_{0,0} - C_{0,0}^t) = 0$, we see that

$$\sum l_k \deg \mathbf{p}_k \equiv \sum n_j \pmod{2}.$$

Since $\text{sgn}(\Gamma(\partial[b, \mathbf{p}, 0])^{q-1}) = (-1)^{\deg \mathbf{p}}$, we get

$$\text{sgn}(\Gamma(\partial B_{1,0})^{q-1}) = (-1)^{\sum l_k \deg \mathbf{p}_k} = (-1)^{\sum n_j} = \theta_t.$$

So we need to relate the sign of $\Gamma(\partial B_{1,0})^{q-1}$ to $v/\sigma_t v$, for which we consider two cases separately.

First Case: \mathbf{a} represents the basis class $[0, \mathbf{p}q, -2]$. Write $d_p = \deg \mathbf{p}$ and $N_p = \#\{\mathbf{a} : \text{monic}, d_a < d_p, \text{sgn}(\{\frac{t\mathbf{a}}{\mathbf{p}}\}) \notin \mathbb{F}_q^{*2}\}$. Using Theorem 8 and the analogue of the classical Gauss lemma, we get by direct calculation

$$\frac{v}{\sigma v} = (-1)^{d_p N_q + d_q N_p} = \text{sgn}(\Gamma(\partial B_{1,0})^{q-1}).$$

Thus we have the result in the first case.

Second Case: \mathbf{a} represents a canonical basis class indexed by a squarefree polynomial \mathbf{g} divisible by at least four distinct irreducibles. In this case $v \in K$ by Theorem 8, and so $\frac{v}{\sigma v} = 1$. We define some operators on \mathbb{SK} as follows.

$$\begin{aligned} t &: [a, *, *] \mapsto [a^t, *, *], \\ I &: [a, *, k] \mapsto \langle a \rangle [a, *, k-1] \quad \text{for } k \text{ odd}, \\ J &: [a, *, k] \mapsto \sum_{l=0}^{\kappa(a)-1} [\gamma^{l-\kappa(a)} a, *, k-1] \quad \text{for } k \text{ even}, \end{aligned}$$

where $\kappa(a)$ is defined by $\text{sgn}(a) = \gamma^{\kappa(a)}$ with $0 \leq \kappa(a) < q-1$. Note that $JI = 0$.

Let C be the cycle obtained by the semi-canonical lifting of k_g . Our aim is to compute $\text{sgn}(\Gamma(\partial B_{1,0})^{q-1}) = (-1)^{\sum l_k \deg \mathbf{p}_k}$, where $B_{1,0} = \sum l_k [c_k, \mathbf{p}_k, 0]$. So we only need to consider the parities of the total sum of $B_{1,0}^{\{\mathbf{p}_k\}}$. Let $C_{n,-n} = [0, \mathbf{g}, -n]$. Then we have

$$B_{1,0} = (J\partial It - JtI\partial)C_{2,-3} + J\partial IJ\partial E + J\partial I\partial JF,$$

for some chains $E, F \in \mathbb{SK}$. A straightforward but tedious computation shows that $TS(B_{1,0}^{\{\mathbf{p}\}}) \equiv 0 \pmod{q-1}$. Then $\text{sgn}(\Gamma(\partial B_{1,0})^{q-1}) = 1$, which implies the result. Lemma 12 and Proposition 14 are used in the course of the computation. \square

We know from Proposition 9 that $\mathbb{FK}_f(\sqrt[q-1]{e(\mathbf{a})})$ is a Galois extension of K , where \mathbb{F} is the quadratic extension of \mathbb{F}_q . But in the second case we get more.

Theorem 16. *Let $\mathbf{a} \in H^2(J, \mathbb{U})$ represent a canonical basis class indexed by a squarefree polynomial \mathbf{f} divisible by at least four distinct irreducibles. Then $K_f(\sqrt[q-1]{e(\mathbf{a})})$ is Galois over K .*

Proof. Since $\frac{e(\mathbf{a})}{\sigma_t e(\mathbf{a})} = \frac{e(\mathbf{a})}{e(\mathbf{a}^t)} = \theta_t e(\mathbf{b})^{q-1}$, for some \mathbf{b} , $K_f(\sqrt[q-1]{e(\mathbf{a})})$ is Galois over K if and only if $\theta_t = 1$. Thus the result follows from the proof of Theorem 15. \square

We have seen in the proof of Theorem 15 that if \mathbf{a} represents a canonical basis class indexed by a square-free polynomial \mathbf{g} divisible by at least four irreducibles, then the total degree of $B_{1,0}^{\{\mathfrak{p}\}}$ is divisible by $q-1$ for any prime \mathfrak{p} and $\theta_t = 1$. Hence $\Gamma(\partial B_{1,0}) \in K^*$. From the proof of Proposition 9, we have

$$\left(\frac{\Gamma(\mathbf{a})}{\Gamma(\mathbf{a}^t)}\right)^{q-1} = \Gamma(\partial B_{1,0})^{q-1}\Gamma(\delta B_{0,1})^{q-1} = \Gamma(\partial B_{1,0})^{q-1}r(B_{0,1})^{1-q}\theta_t\frac{e(\mathbf{a})}{e(\mathbf{a}^t)}.$$

Finally, we have the following stronger version of Theorem 8.

Theorem 8’. *Let $n \geq 4$ be an even positive integer. Let $C = \bigoplus C_{i,-i}$ be the semi-canonically lifted cycle from the basis class $[0, \mathbf{g}, -n]$, where \mathbf{g} is a square-free monic polynomial divisible by n irreducible polynomials. Let $\mathbf{a} = \sum m_i[a_i]$, where $C_{0,0} = \sum m_i[a_i, 1, 0]$. Then*

$$\Gamma(\mathbf{a})^{q-1} = re(\mathbf{a}) \quad \text{and} \quad \Gamma(\mathbf{a}^t)^{q-1} = rs^{q-1}e(\mathbf{a}^t)$$

for some $r, s \in K^*$.

In [BY] we proved that if $\mathbf{a} \in H^2(J, \mathbb{U})$ represents the basis class $[0, \mathfrak{p}\mathbf{q}, -2]$, then $K_{\mathfrak{p}\mathbf{q}}(\sqrt[q-1]{e(\mathbf{a})})$ is nonabelian over K . Here we also give an example where $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))$ is not abelian over K .

Example. We assume that $q = 3$. We can easily compute that the cycle $C = C_{0,0} \oplus C_{1,-1} \oplus C_{2,-2}$ is the semi-canonically lifted cycle of the canonical basis class $[0, T(T+1), -2]$, where

$$\begin{aligned} C_{2,-2} &= [0, T(T+1), -2], \\ C_{1,-1} &= [\frac{1}{T}, T+1, -1] - [\frac{1}{T+1}, T, -1], \\ C_{0,0} &= [\frac{1}{T+1}, 1, 0] - [\frac{T-1}{T(T+1)}, 1, 0]. \end{aligned}$$

Thus

$$\mathbf{a}_{T(T+1)} = [\frac{1}{T+1}] - [\frac{T-1}{T(T+1)}].$$

A simple computation gives

$$\Gamma(\mathbf{a}_{T(T+1)})^2 = \sqrt{\frac{T}{T+1}} \frac{e(\frac{\tilde{\pi}}{T+1})}{e(\frac{(T-1)\tilde{\pi}}{T(T+1)})} = u,$$

using the relation $\Gamma(\mathbf{a}_{T(T+1)})^2 = \Gamma(\delta SC_{0,0})\Gamma(\partial SC_{1,-1})$. Let $\sigma = \sigma_{T-1}$ and $\tau = \sigma_{-T+1}$. Let $\lambda = e(\frac{\tilde{\pi}}{T(T+1)})$. Then we can check that

$$\frac{u}{\sigma u} = \lambda^2, \quad \frac{u}{\tau u} = \lambda^2, \quad \text{and} \quad \frac{u}{\sigma \tau u} = 1.$$

Let

$$v_\sigma = \lambda \quad v_\tau = \lambda \quad \text{and} \quad v_{\sigma\tau} = 1.$$

Let $\tilde{\eta} \in \text{Gal}(K_{\mathfrak{f}}(\Gamma(\mathbf{a}))/K)$ be the lifting of $\eta \in \text{Gal}(K_{\mathfrak{f}}/K)$ such that $v_\eta \tilde{\eta} \sqrt{u} = \sqrt{u}$. Then using the fact that $\lambda^4 + (T+1)\lambda^2 + 1 = 0$, we get $\tilde{\sigma}\tilde{\tau} = -\tilde{\tau}\tilde{\sigma}$ on \sqrt{u} .

Remark. It would be very interesting to know whether $K_{\mathfrak{f}}(\Gamma(\mathbf{a}))$ is abelian over K , or equivalently by the last theorem, whether $K_{\mathfrak{f}}(\sqrt[q-1]{e(\mathbf{a})})$ is abelian over K , if \mathbf{a} represents a canonical basis class indexed by a monic square-free polynomial divisible by at least four irreducibles. In the classical case this is verified by Das

([Da], Theorem 21), with the aid of a theorem of Deligne (Theorem 7.18(b) of [De], Theorem 19 of [Da]), that is,

$$(4.1) \quad \sigma \left(\frac{\Gamma(\mathbf{a})}{\tau\Gamma(\mathbf{a})} \right) = \frac{\Gamma(\mathbf{a}^t)}{\tau\Gamma(\mathbf{a}^t)},$$

where $\mathbf{a} \in H^2(J, \mathbb{U})$, $\sigma, \tau \in \text{Gal}(\bar{K}/K)$ and the restriction of σ on K_f is σ_t . Here \bar{K} is the separable closure of K . Note that it is easy to see that $\frac{\Gamma(\mathbf{a})}{\tau\Gamma(\mathbf{a})} \in K_f$ by Theorem 8. We also note that when \mathbf{a} represents a semi-canonically lifted cycle from a basis class, then

$$\sigma \left(\frac{\Gamma(\mathbf{a})}{\tau\Gamma(\mathbf{a})} \right)^{q-1} = \left(\frac{\Gamma(\mathbf{a}^t)}{\tau\Gamma(\mathbf{a}^t)} \right)^{q-1},$$

using Theorem 8'.

If one disposes of an analogue of Deligne's theorem above, then one can easily show, with the aid of Theorem 8' and following the same method as in [Da], that $K_f(\Gamma(\mathbf{a}))$ is abelian over K if \mathbf{a} satisfies the above conditions.

Sinha [Si] has proven Deligne's reciprocity for function fields (Theorem 7.18(a) of [De]) using Anderson's theory of solitons. Thus one may also use the theory of solitons to prove the analogue of Deligne's theorem (Theorem 7.18(b) of [De]), which is beyond the reach of the present paper. However, we hope that an elementary proof using the double complex may be possible. Anderson's recent work on the epsilon extension yields an elementary method to show that $K(\sqrt[q-1]{e(\mathbf{a})})$ is abelian over K .

Let K^{ab} be the maximal abelian extension of K and let $G^{ab} = \text{Gal}(K^{ab}/K)$. Following Anderson [An2], two of us [BY] defined an injective homomorphism

$$\mathbf{D} : H^0(G^{ab}, K^{ab*}/K^{ab*(q-1)}) \longrightarrow \bigwedge^2 H^1(G^{ab}, \mathbb{Z}/(q-1)\mathbb{Z})$$

and showed it is an isomorphism. We can also express \mathbf{D} explicitly; see [BY, Sect. 3.5] for detail. The map \mathbf{D} has the property that for $u \in K^{ab*}$,

$$\mathbf{D}(u \bmod K^{ab*(q-1)}) = 0 \iff \sqrt[q-1]{u} \in K^{ab}.$$

This proposes an elementary method for showing that $\sqrt[q-1]{e(\mathbf{a})} \in K^{ab}$ if \mathbf{a} satisfies the condition above. But the calculation of $\mathbf{D}(\sqrt[q-1]{e(\mathbf{a})})$ would be too complicated to take. About this question, we refer the reader to Remark 4.4.2 in [An2].

ACKNOWLEDGEMENTS

We are very grateful to the referee for many valuable suggestions, especially, for the definition of the vertical shift operator, which is more effective than our old one and enabled a significant shortening of the paper.

REFERENCES

- [An1] G. Anderson, *A double complex for computing the sign-cohomology of the universal ordinary distribution*, Contemp. Math. **224** (1999), 1-27. MR **99k**:11169
- [An2] G. Anderson, *Kronecker-Weber plus epsilon*, Duke Math. J. **114** (2002), 439-475.
- [BGY] S. Bae, E.-U. Gekeler, and L. Yin, *Distributions and Γ -monomials*, Math. Ann. **321** (2001), 463-478. MR **2002i**:33002
- [BY] S. Bae, and L. Yin, *Carlitz-Hayes plus Anderson's epsilon*, Submitted for publication.
- [Da] P. Das, *Algebraic gamma monomials and double coverings of cyclotomic fields*, Trans. Amer. Math. Soc. **352** (2000), 3557-3594. MR **2000m**:11107

- [De] P. Deligne, J. Milne, A. Ogus, and K. Shih, *Hodge cycles, Motives, and Shimura Varieties*, Lecture Notes in Math. **900** (1982). MR **84m**:14046
- [GR] S. Galovich and M. Rosen, *Distributions on rational function fields*, Math. Ann. **256** (1981), 549-560. MR **83e**:12007
- [Ha] D. Hayes, *Explicit class field theory for rational function fields*, Trans. Amer. Math. Soc. **189** (1974), 77-91. MR **48**:8444
- [Ou] Y. Ouyang, *The group cohomology of the universal ordinary distribution*, J. reine angew. Math. **537** (2001), 1-32. MR **2002f**:11148
- [Si] S. Sinha, *Deligne's reciprocity for function fields*, J. Number Theory **63** (1997), 65-88. MR **98a**:11074
- [Th] D. Thakur, *Gamma functions for function fields and Drinfeld modules*, Ann. of Math. **134** (1991), 25-64. MR **92g**:11058

DEPARTMENT OF MATHEMATICS, KAIST, TAEJON 305-701, KOREA

E-mail address: shbae@math.kaist.ac.kr

DEPARTMENT OF MATHEMATICS, SAARLAND UNIVERSITY, D-66041 SAARBRUCKEN, GERMANY

E-mail address: gekeler@math.uni-sb.de

DEPARTMENT OF MATHEMATICS, CHUNGNAM NATIONAL UNIVERSITY, TAEJON 305-764, KOREA

E-mail address: plkang@math.cnu.ac.kr

DEPARTMENT OF MATHEMATICAL SCIENCES, TSINGHUA UNIVERSITY, BEIJING 100084, PEOPLE'S REPUBLIC OF CHINA

E-mail address: lsyin@math.tsinghua.edu.cn

REMARKS ABOUT UNIFORM BOUNDEDNESS OF RATIONAL POINTS OVER FUNCTION FIELDS

LUCIA CAPORASO

ABSTRACT. We prove certain uniform versions of the Mordell Conjecture and of the Shafarevich Conjecture for curves over function fields and their rational points.

1. INTRODUCTION AND PRELIMINARIES

A curve X of genus at least 2 defined over a function field L has only finitely many L rational points, unless it is isotrivial. Similarly, a curve of genus at least 2 defined over a number field F has a finite set of F -rational points. These well-known facts are celebrated theorems of Y. Manin and G. Faltings, originally conjectured by L. J. Mordell and S. Lang.

We study here questions of uniformity for the cardinality of such sets of rational points, in the function field case. For number fields, there are a number of open conjectures, such as the following (Uniform Mordell Conjecture for number fields): *Fix $g \geq 2$ and a number field F ; there exists a number $B_g(F)$ such that any curve of genus g defined over F has at most $B_g(F)$ rational points over F .* Interest in such problems was revived after it was proved in [CHM] that the conjecture above is a consequence of a famous, open, conjecture (usually attributed to S. Lang and E. Bombieri) on the non-density of rational points in varieties of general type (see also [Ab], [AV] and [Pa]).

In this paper we investigate similar issues for curves over function fields. Some partial results were obtained in [Mi] and in [C] where the existence of uniform bounds for the sets of rational points is established. Such bounds depend on suitable numerical invariants of the function field, on the genus g of the curves and on the degree of the locus of bad reduction (that is, the locus of singular fibers).

We shall also study here the strictly related “uniform Shafarevich problem”; a famous theorem of A. N. Parshin and S. Ju. Arakelov ([Ar] and [P]) states that *if B is a smooth complex curve and $S \subset B$ a finite subset, then there exists only a finite number of non-isotrivial families of smooth curves of fixed genus $g \geq 2$ over $B - S$.* Parshin first proved it under the assumption that $S = \emptyset$; Arakelov generalized it a few years later. In [P] Parshin shows also that the above theorem implies finiteness of rational points for non-isotrivial curves of genus at least 2, providing the above mentioned link between the Shafarevich problem and the Mordell problem. Recall that his argument, known as the “Parshin trick”, is valid for both number fields and function fields.

Received by the editors January 10, 2001 and, in revised form, September 24, 2001.
 2000 *Mathematics Subject Classification*. Primary 14H05, 14H10.

A first uniform version of the theorem of Parshin and Arakelov above is obtained in [C]. We here generalize it by a stronger uniform result valid for families of curves over bases of any dimension. This is done in Section 2, where we obtain bounds (for the sets of curves with fixed degeneracy locus as well as for the sets of rational points) that only depend on the degree of a polarization on the base variety, and on the degree of the locus of bad reduction. A stronger result can be obtained for curves having good reduction in codimension 1 (Theorem 3). In Section 3 we will consider families with maximal variation of moduli, using the geometry of the moduli space of curves to approach our problems.

We work over \mathbb{C} ; by V we shall denote a smooth, irreducible, projective variety over \mathbb{C} , whose field of rational functions will be $L := \mathbb{C}(V)$. Special interest will be given to varieties of dimension 1, for which we shall use the following notation: B is a smooth irreducible curve and K its field of rational functions. We fix integers $q \geq 0$, $g \geq 2$ and $s \geq 0$ throughout. The genus of B will be denoted by q .

We shall consider smooth curves of genus g over the function field L (or K), which can also be viewed as families of curves over V , such that there is a nonempty open subset of V over which the fibers are all smooth. We shall always assume that such a family (or curve) is not isotrivial, i.e., the smooth fibers are not all isomorphic.

To be more precise, we introduce the following sets: let B be a fixed curve and let $S \subset B$ be a finite set of points.

Definition. $F_g(B, S)$ shall denote the set of equivalence classes of non-isotrivial families $f : X \rightarrow B$ such that X is a smooth relatively minimal surface and the fiber X_b over every $b \notin S$ is a smooth curve of genus g . Two such families $f_i : X_i \rightarrow B$ for $i = 1, 2$ are equivalent if there is a commutative diagram

$$\begin{array}{ccc} X_1 & \xrightarrow{\alpha'} & X_2 \\ f_1 \downarrow & & \downarrow f_2 \\ B & \xrightarrow{\alpha} & B \end{array}$$

where the two horizontal arrows are birational maps.

Using a different terminology, $F_g(B, S)$ is the set of K -isomorphism classes of non-isotrivial curves of genus g over K , having good reduction outside of S . The theorem of Parshin and Arakelov says that $F_g(B, S)$ is finite. Theorem 3.1 of [C] states that *there exists a number $P(g, q, s)$ such that $|F_g(B, S)| \leq P(g, q, s)$ for every curve B of genus q and for every subset S having at most s points*. We show here (in the end of Section 2) that this result is sharp in the sense that such a bound must depend on s .

We are interested in function fields of higher transcendence degree. We can generalize the definition of $F_g(B, S)$ as follows. Let $T \subset V$ be a closed subscheme.

Definition. $F_g(V, T)$ shall be the set of equivalence classes of non-isotrivial families of smooth curves of genus g over $V - T$ (the equivalence relation is the same as above, with B replaced by $V - T$).

By the existence and unicity of minimal models for smooth surfaces, this definition coincides with the previous one if $\dim V = 1$. It follows from the results in [C] (3.4) that $F_g(V, T)$ is finite. Our best result on $F_g(V, T)$ is Theorem 1.

If X is a curve defined over a field L , we shall denote by $X(L)$ the set of its L -rational points. If X has genus at least 2 and it is not isotrivial, the theorem of Manin says that $X(L)$ is finite. Consider now the Uniformity Conjecture for

rational points over function fields, which can be stated as its arithmetic analogue: Let L be a function field over \mathbb{C} and let $g \geq 2$ be an integer. There exists a number $N_g(L)$ such that for every non-isotrivial curve X of genus g defined over L we have $|X(L)| \leq N_g(L)$. For results relating it to the Lang Conjectures about the distributions of rational points on varieties of general type, see the work of D. Abramovich and J. F. Voloch [AV].

Such a conjecture remains open; our results in that direction are Theorems 2 and 3 and Proposition 4.

A final piece of notation. M_g denotes the moduli variety of smooth curves of genus g and \overline{M}_g its compactification via Deligne-Mumford stable curves. They are both integral, normal varieties of dimension $3g - 3$. A universal curve exists only on a proper open subset of M_g (and of \overline{M}_g). In particular, a morphism $\phi : Z \rightarrow M_g$ does not necessarily come from a family of curves over Z . If this is the case, that is, if there exists a family of smooth curves $X \rightarrow Z$ such that for every $z \in Z$, $\phi(z)$ is the isomorphism class of the fiber of X over z , we shall say that ϕ is a *moduli map*.

2. UNIFORMITY RESULTS FOR FUNCTION FIELDS OF HIGH TRANSCENDENCE DEGREE

We start by a uniform generalization of the theorem of Parshin and Arakelov. The result below is a strengthening of 3.4 and 3.5 in [C]; in fact, the bound H here is independent of the dimension of V and of r . Such an improvement is obtained by a small technical modification of the methods in [C].

Notice that the statement below remains true if V is replaced by an integral, possibly singular, projective variety. The proof is essentially the same.

Theorem 1. *Let $g \geq 2$, $d \geq 1$, $s \geq 0$ be fixed integers. There exists a number $H(g, d, s)$ such that for any smooth, irreducible variety $V \subset \mathbb{P}^r$ of degree d , for any closed subscheme $T \subset V$ of degree s , we have $|F_g(V, T)| \leq H(g, d, s)$. Moreover, if T has codimension at least 2 in V , then the bound H does not depend on s .*

Proof. Step 1: Slicing V into curves of bounded genus. Considering one-dimensional hyperplane sections of V , we see that V can be covered by smooth curves of degree d passing through any of its points; it is a well-known fact that the genus of a curve of degree d in projective space is at most equal to $\binom{d-1}{2}$: just project the curve birationally onto a curve of degree d in \mathbb{P}^2 . Let

$$q = q(d) = \binom{d-1}{2}$$

so that V is covered by curves of geometric genus at most q .

Step 2: Uniform boundedness of moduli maps. By Theorem 3.1 in [C], for any fixed g, q' , and s' there exists a number $P(g, q', s')$ such that for any smooth curve B of genus q' , for any subset $S \subset B$ of at most s' points, we have that $|F_g(B, S)| \leq P(g, q', s')$.

Define

$$H' = \max_{q' \leq q, s' \leq s} P(g, q', s')$$

so that H' only depends on g, d, s ; let $U = V - T$.

We claim that U has at most H' moduli maps to M_g ; that is, we claim that there exist at most H' non-constant, (regular) morphisms $\phi : U \rightarrow M_g$ such that there exists a (not necessarily unique, see below) family of smooth curves over U whose moduli map is ϕ . By contradiction, let $n > H'$ and let us assume that there exist ϕ_1, \dots, ϕ_n distinct such moduli maps $\phi_i : U \rightarrow M_g$. Let $X_i \rightarrow U$ be a non-isotrivial family of smooth curves corresponding to ϕ_i (since ϕ_i is a moduli map, such a family exists, but it is not necessarily unique). Let $U' \subset U$ be the nonempty open subset where $\phi_i(u) \neq \phi_j(u)$ for every $u \in U'$ and for every pair of distinct i, j . Let $p \in U'$ and let $F_i = \phi_i^{-1}(\phi_i(p))$; since ϕ_i is not constant, its fiber F_i through p is a proper closed subset of U' . Therefore, there exists a curve $B \subset V$ of genus at $q' \leq q$ such that $p \in B$ and such that $B \not\subset F_i$ for every $i = 1, \dots, n$; thus the restriction of X_i to B is not isotrivial for every i . Let $S = (B \cap T)_{\text{red}}$. Let $Y_i \rightarrow B$ be the smooth relatively minimal completion over B of the restriction of X_i to B . By construction, Y_1, \dots, Y_n are different elements of $F_g(B, S)$, which is a contradiction, since $F_g(B, S)$ has at most $P(g, q', s') \leq H' < n$ elements. This proves the claim. Notice that if T has codimension at least 2 in V we can always choose our B so that it does not intersect T at all, and hence S can be taken to be the empty set and H' does not depend on s .

Conclusion. Given a moduli map $\phi : U \rightarrow M_g$ the set of families that have ϕ as moduli map is uniformly bounded; in fact, it is bounded above by a function of g only (see [C], Lemma 3.3); hence we are done. \square

A similar argument yields the following uniformity statement for rational points, stronger than 4.3 and 4.4 in [C]:

Theorem 2. *Let $g \geq 2$, $d \geq 1$, $s \geq 0$ be fixed integers. There exists a number $N(g, d, s)$ such that for any smooth, irreducible variety $V \subset \mathbb{P}^r$ of degree d , for any closed subscheme $T \subset V$ of degree s and for any non-isotrivial curve X of genus g defined over $L = \mathbb{C}(V)$ and having good reduction outside of T , we have $|X(L)| \leq N(g, d, s)$. Moreover, if T has codimension at least 2 in V , then the bound N does not depend on s .*

Proof. Step 1: Repeat word for word Step 1 in the proof of the previous theorem.

Step 2: Theorem 4.2 in [C] says that if g, q' , and s' are fixed nonnegative integers, there exists a number $M(g, q', s')$ such that for any curve B of genus q' , for any subset S of at most s' points in B , and for any curve $X_B \in F_g(B, S)$ we have that

$$|X_B(\mathbb{C}(B))| \leq M(g, q', s').$$

Arguing as in the proof of 4.4 of [C] one gets that defining

$$N(g, d, s) := \max_{q' \leq q, s' \leq s} M(g, q', s')$$

will suffice for our statement. \square

To conclude, we show that for curves having good reduction in codimension 1, stronger finiteness results hold. Let L be a function field over \mathbb{C} and let V be a smooth, projective, complex variety of positive dimension such that $L = \mathbb{C}(V)$.

Definition. Let $C_g^2(L)$ be the set of L -isomorphism classes of non-isotrivial curves of genus g over L having good reduction in codimension 1.

In other words, $C_g^2(L)$ is the set of equivalence classes of non-isotrivial families $X \rightarrow V$ of curves of genus g over V such that there exists a closed subscheme

$T \subset V$ of codimension at least 2 with the property that X_v is smooth for every $v \notin T$.

Theorem 3.

- a) $C_g^2(L)$ is finite.
- b) There exists a number $N_g^2(L)$ such that for every curve $X \in C_g^2(L)$ we have $|X(L)| \leq N_g^2(L)$.

Proof. We shall use moduli maps. Denote by $M_g^2(L)$ the set of equivalence classes of non-constant rational maps $\phi : V \rightarrow M_g$ such that there exists an open subset $U^\phi \subset V$ with the following properties:

1. The complement of U^ϕ has codimension at least 2 in V .
2. ϕ is regular on U^ϕ .
3. There exists a (non-isotrivial) family of smooth curves of genus g over U^ϕ such that ϕ is its moduli map.
4. Two such maps ϕ and ψ are equivalent iff they coincide on some (nonempty) open subset of V .

There is a natural surjective map of sets:

$$\mu : C_g^2(L) \longrightarrow M_g^2(L)$$

sending a curve over L to its moduli map (it is easy to see that μ is well defined). Now, μ has finite fibers (Lemma 3.3 in [C]) and is surjective by definition. Thus $C_g^2(L)$ is finite if and only if $M_g^2(L)$ is finite.

Part b) is an immediate consequence of part a), by the theorem of Manin. We will prove our result by showing that $M_g^2(L)$ is finite by induction on $\dim V$. If $\dim V = 1$, then the finiteness of $C_g^2(L)$ and of $M_g^2(L)$ is the theorem of Parshin (the locus of bad reduction being empty in such a case). Then let $\dim V \geq 2$ and suppose that $M_g^2(L)$ is infinite. Notice that $M_g^2(L)$ is dominated by a union of finite sets as follows: if T is a closed subset of V , denote by $M_g(V, T)$ the set of equivalence classes of moduli maps to M_g that are regular on $V - T$; then $M_g(V, T)$ is finite, by Theorem 1 and Lemma 3.3 in [C]. We have a natural, surjective map

$$\bigcup_{\text{codim}_V T \geq 2} M_g(V, T) \longrightarrow M_g^2(L);$$

hence, if $M_g^2(L)$ is infinite, so is the union on the left-hand side. Then there exists a countable collection $\{T_n, n \in \mathbb{Z}\}$, with T_n a closed subset of V of codimension at least 2, such that the set

$$M := \bigcup_{n \in \mathbb{Z}} M_g(V, T_n)$$

is infinite. Now, M itself being a countable set, we shall put an ordering on it:

$$M = \{\phi^i, i \in \mathbb{N}\}.$$

For every pair of distinct i, j , denote by $U^{i,j}$ the nonempty open subset of V such that $U^{i,j} \subset U^{\phi^i} \cap U^{\phi^j}$ and $\phi^i(u) \neq \phi^j(u)$ for every $u \in U^{i,j}$. The $U^{i,j}$ s form a countable collection of nonempty open subsets of V , whose intersection I is dense in V . Let $p \in I$ and, for every $i \in \mathbb{N}$, let $F_i = (\phi^i)^{-1}\phi^i(p)$ be the fiber of ϕ^i through p . Since ϕ^i is non-constant (by assumption), F_i is a proper closed subset of V . Thus, the complement of $\bigcup_{i \in \mathbb{N}} F_i$ intersects I in a subset J , with J dense in V . Fix a (non-degenerate) projective model of V in some projective space. Then there

exists a hyperplane H such that $p \in H$, such that $H \cap J \neq \emptyset$ and such that H does not contain any T_n . Letting $W = H \cap V$, we can furthermore choose H so that W is smooth. By construction we have

- (a) $\dim W = \dim V - 1$;
- (b) $\dim T_n \cap W = \dim T_n - 1 \leq \dim W - 2$
(since H does not contain any T_n);
- (c) $\forall \phi^i \in M$, the restriction $\phi^i|_W$ is not constant
(since $p \in W$ and $W \cap J \neq \emptyset$);
- (d) $\forall i \neq j$ we have $\phi^i|_W \neq \phi^j|_W$
(since $W \cap I \neq \emptyset$);

hence $\phi^i|_W \in M_g^2(\mathbb{C}(W))$ and the restriction to W gives an inclusion (by (d) above) $M \hookrightarrow M_g^2(\mathbb{C}(W))$. Thus $M_g^2(\mathbb{C}(W))$ is infinite. This is a contradiction with the inductive assumption. \square

See [Md] for an analogue over \mathbb{Q} . Part a) of this result should be compared with the examples of A. Beauville (in [B], section 5) or with the example below. They show that the assumption that the curves have good reduction in codimension 1 is crucial; that is, a) is false without that assumption. On a different vein, compare also with Proposition 4. The example that we are going to describe shows that there is no hope of getting a substantially stronger uniform version of the Shafarevich Conjecture for function fields; in other words, any uniform bound on $|F_g(B, S)|$ must depend on the degree of S .

What happens to the cardinality of $F_g(B, S)$ when s grows while g and q (or even B) stay fixed? The way we defined $F_g(B, S)$, it is an exercise to show that its cardinality is not bounded; but this is just because the families parametrized by $F_g(B, S)$ are not required to have a singular fibers over S . The interesting question is about the asymptotics of the cardinality of that subset of $F_g(B, S)$ parametrizing families of curves that have a singular fiber over every point of S . We will make this precise now, describing an example suggested by J. de Jong, showing that the set of fibrations with fixed degeneracy locus is not bounded, as the cardinality of the degeneracy locus grows.

Fix $g \geq 2$ and $B = \mathbb{P}^1$. Given a subset $S \subset \mathbb{P}^1$ denote by $F(S) \subset F_g(\mathbb{P}^1, S)$ the set of all genus g non-isotrivial fibrations $X \rightarrow \mathbb{P}^1$ such that the fiber X_b is smooth if and only if $b \notin S$.

Let $S = \{a_1, \dots, a_s\}$ be a set of generic points in \mathbb{P}^1 , and let $I \cup J = \{1, 2, \dots, s\}$ be a partition of $\{1, 2, \dots, s\}$ in two disjoint subsets such that $|I| = 5$. Define a non-isotrivial fibration X_I of curves of genus 2 over \mathbb{P}^1 by the affine equation

$$y^2 = (x - t)\prod_{i \in I}(x - a_i)\prod_{j \in J}(t - a_j)$$

with t an affine coordinate in \mathbb{P}^1 . For $t \notin S$ (and $t \neq \infty$) we get a smooth curve of genus 2. For $t = a_i$ with $i \in I$, we get a nodal curve and for $t = a_j$, $j \in J$, we get a singular, non-reduced curve. Thus $X_I \in F(S \cup \infty)$ and, by varying the partition $I \cup J$, we get a total of $\binom{s}{5}$ different such fibrations. Hence the cardinality of $F(S)$ goes to infinity, as $|S|$ grows.

One final word about this example.

First, we make two comments: the given family has fibers of genus 2, but of course one can construct the same example for any genus (just replace the integer 5 by a higher odd number), obtaining families of hyperelliptic curves.

The second comment is about those singular fibers over a_j with $j \in J$ that are not stable curves; their semistable reduction is actually a smooth curve. The remaining 5 fibers over a_i are instead nodal. In other words, the moduli map ϕ_I associated to the family $X_I \rightarrow \mathbb{P}^1$,

$$\phi_I : \mathbb{P}^1 \rightarrow \overline{M}_2$$

(such that $\phi_I(t)$ is the isomorphism class of the fiber of X_I over t) intersects the boundary Δ_2 in exactly 5 points, regardless of the cardinality of S .

We ask:

- (a) Can one find similar examples whose fibers do not belong to any proper closed subset of M_g ?
- (b) Is the same “unboundedness” result true for families of stable curves? In other words, does there exist a similar example all of whose singular fibers are nodal?

3. UNIFORMITY FOR “TRULY VARYING” CURVES

This section contains results that are independent of the degeneracy locus. Given a family $X \rightarrow V$ of generically smooth curves of genus g over V , we get a natural rational map $\phi : V \rightarrow M_g$ (regular on a nonempty open subset of V). The dimension of the image of ϕ is called the *variation of moduli* of the family; we shall say that the family has *maximal variation of moduli* if

$$\dim \operatorname{Im} \phi = \min\{\dim V, 3g - 3\}.$$

We shall say that a curve over $L = \mathbb{C}(V)$ has maximal variation of moduli if a corresponding family of curves over V does.

Thus the condition of having maximal variation of moduli can be interpreted as saying that the family (or the curve) is *truly varying* and can be viewed as a generalization of the non-isotriviality condition. Obviously, if the base field has transcendence degree 1, a curve is non-isotrivial if and only if it has maximal variation of moduli.

Definition. Let L be a function field. We define $C_g(L)$ to be the set of L -isomorphism classes of curves of genus g defined over L and having maximal variation of moduli.

Proposition 4. *Let $g \geq 24$ and let L be a function field of transcendence degree $3g - 3$. Then*

- a) $C_g(L)$ is finite.
- b) There exists a number $N(L, g)$ such that for every curve X of genus g defined over L and having maximal variation of moduli, we have $|X(L)| \leq N(L, g)$.
- c) There exists a function $P_g(n, m)$ such that for every V of general type, we have $|C_g(L)| \leq P_g(\dim V, K_V^{\dim V})$.

Proof. The assumption $g \geq 24$ implies that \overline{M}_g is of general type (for this famous result of J. Harris and D. Mumford we refer to [HMu] and to 6F in [HM]).

Denote by $R(V, \overline{M}_g)$ the set of dominant, rational maps from V to \overline{M}_g . A theorem of Kobayashi and Ochiai [KO] implies that, \overline{M}_g being of general type, $R(V, \overline{M}_g)$ is finite. Notice now that there is a natural bijection between $C_g(L)$ and $R(V, \overline{M}_g)$: to a truly varying curve X of genus g over L we can associate its moduli map $\phi_X \in R(V, \overline{M}_g)$. The fact that such a correspondence is bijective follows from the existence of the universal curve over an open subset of \overline{M}_g . Thus $C_g(L)$ is finite.

By the theorem of Manin, any curve in $C_g(L)$ has a finite set of L -rational points. Thus part b) follows immediately from a).

Part c) is proved like part a); we can in this case apply a strengthening of the theorem of Kobayashi and Ochiai provided by T. Bandman and D. Markushevich. From [BM] we obtain that, V and \overline{M}_g being of general type and \overline{M}_g having canonical singularities (Theorem 1 in [HMu]), there exists a function of g , of $\dim V$ and of $K_V^{\dim V}$ bounding the cardinality of $R(V, \overline{M}_g)$ and hence that of $C_g(L)$. \square

Let $u : \mathcal{C}_g \rightarrow M_g^\circ$ be the universal curve over the moduli space of automorphism free smooth curves of genus g , so that the fiber of u over the point corresponding to the curve X is X itself.

It is a well-known fact (see [HM], 2D) that u has no rational sections; thus, \mathcal{C}_g has no rational point over the function field of M_g . In fact, much more is known: the Picard group of \mathcal{C}_g is generated over the Picard group of M_g by the relative dualizing sheaf ω_u ; therefore, a multisection of u must have degree over M_g° equal to a multiple of $2g - 2$.

We apply this to obtain that if V is a variety of dimension $3g - 3$ and X is a curve of genus g over L having maximal variation of moduli, then a necessary condition for X to have a rational point over L is that its moduli map have degree equal to a multiple of $2g - 2$. This follows easily by looking at the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\gamma} & \mathcal{C}_g \\ f \downarrow \uparrow \sigma & & \downarrow u \\ V & \xrightarrow{\phi_X} & M_g \end{array}$$

where the horizontal arrows are rational maps and σ is the rational section corresponding to a rational point of X over L . Let $\tau = \gamma \circ \sigma : V \rightarrow \mathcal{C}_g$ and let $\rho : \text{Im } \tau \rightarrow M_g$; by what we said, $\deg \rho = n(2g - 2)$ for some integer n . We finally obtain

$$\deg \phi_X = \deg \tau \cdot \deg \rho = m(2g - 2),$$

where by $\deg \phi_X$ we mean the degree of the restriction of ϕ_X to the nonempty open subset of V and where ϕ_X is a regular and finite map. Let us call such a number $\deg \phi_X$ the *modular degree* of a family $X \rightarrow V$; this definition is general, provided that $X \rightarrow V$ has maximal variation of moduli and that $\dim V \leq 3g - 3$. We just proved the following

Lemma 5. *Let V be a variety of dimension $3g - 3$ with function field L and let X be a smooth curve of genus g over L having maximal variation of moduli. Then either $X(L) = \emptyset$ or the modular degree of X is a multiple of $2g - 2$.*

The following well-known conjecture is open:

Geometric Lang Conjecture. *Let W be a variety of general type defined over \mathbb{C} . Then there exists a proper closed subvariety Z_W of W containing all positive-dimensional subvarieties of W that are not of general type.*

In particular, according to such a conjecture, all curves in W having genus at most 1 are contained in Z_W .

Consider now \overline{M}_g , and let $Z_g \subset \overline{M}_g$ be defined as the closure of the union of all integral curves in \overline{M}_g having geometric genus at most equal to 1. Since \overline{M}_g is of general type if $g \geq 24$, the above conjecture would imply that Z_g is a proper, closed subset of \overline{M}_g for all $g \geq 24$.

As a consequence, we get the following.

Lemma 6. *Let $g \geq 24$ and let B be a curve of genus q . The Geometric Lang Conjecture implies that if $X \rightarrow B$ is a non-isotrivial family of curves of genus g , passing through the general point of M_g , then the modular degree of X is at most $q - 1$.*

Proof. As we mentioned above, the union of all curves in \overline{M}_g of genus at most 1 is contained in a proper closed subset Z_g of \overline{M}_g . The condition that the given family of curves goes through the general point of \overline{M}_g , combined with the Geometric Lang Conjecture, implies that $\text{Im } \phi_X \not\subset Z_g$. Thus the geometric genus of $\text{Im } \phi_X$ is at least 2. By the Riemann-Hurwitz formula, the degree d of a dominant map of a curve B of genus q onto a curve C of geometric genus $p \geq 2$ is at most equal to $q - 1$. In fact, the formula gives

$$d = \frac{2q - 2 - r}{2p - 2} \leq \frac{q - 1}{p - 1} \leq q - 1$$

since $r \geq 0$ (being the degree of the ramification divisor) and $p \geq 2$ by assumption. \square

ACKNOWLEDGMENTS

I am grateful to Olivier Debarre and to Johan de Jong for useful conversations and to Felipe Voloch for indicating relevant references. Special thanks to Dan Abramovich for pointing out a serious mistake in a previous version of this paper.

REFERENCES

- [Ab] D. Abramovich: *Uniformity of stably integral points on elliptic curves*, *Inventiones Math.* 127, 307–317 (1997). MR 98d:14033
- [AV] D. Abramovich and J. F. Voloch: *Lang's conjectures, fibered powers, and uniformity*, *New York J. Math.* 2, 20–34 (1996) MR 97e:14031
- [Ar] S. Ju. Arakelov: *Families of algebraic curves with fixed degeneracies*, *Izv. Akad. Nauk. SSSR Ser. Mat.* 35, 1269–1293 (1971). MR 48:298
- [BD] T. Bandman and G. Dethloff: *Estimates of the number of rational mappings from a fixed variety to varieties of general type*, *Ann. Inst. Fourier (Grenoble)* 47 (1997), no. 3, 801–824. MR 98h:14016
- [BM] T. Bandman and D. Markushevich: *On the number of rational maps between varieties of general type*, *J. Math. Sci. Univ. Tokyo* 1 (1994), no. 2, 423–433. MR 96c:14012
- [B] A. Beauville: *Exposé No. 6 in Séminaire sur les pincesaux des courbes de genre au moins deux*, *Astérisque* 86 (1981). MR 83c:14020
- [C] L. Caporaso: *On certain uniformity properties of curves over function fields*, Preprint, AG/9906156, *Compositio Mathematica* 130 (2002), 1–19. MR 2003a:14038

- [CHM] L. Caporaso, J. Harris, and B. Mazur: *Uniformity of rational points*, J. Amer. Math. Soc. 10 (1997), 1–35. MR **97d**:14033
- [DF] M. De Franchis: *Un teorema sulle involuzioni irrazionali*, Rend. Circ. Mat. Palermo 36 (1913), 368.
- [HM] J. Harris and I. Morrison: *Moduli of curves*, Graduate Texts in Mathematics 187, Springer-Verlag, New York, 1998. MR **99g**:14031
- [HMu] J. Harris and D. Mumford: *On the Kodaira dimension of the moduli space of curves*, Inventiones Math. 67 (1982), 23–88. MR **83i**:14018
- [KO] S. Kobayashi and T. Ochiai: *Meromorphic mappings onto compact complex spaces of general type*, Inventiones Math. 31 (1975), 7–16. MR **53**:5948
- [Md] M. Martin-Deschamps: *Conjecture de Shafarevich pour les corps de fonctions sur \mathbb{Q}* , Astérisque No. 127, Appendice à l'exposé IX (1985), 256–259.
- [Ma] Y. Manin: *Rational points of algebraic curves over function fields*, Izv. Akad. Nauk. SSSR Ser. Mat. 27 (1963), 1395–1440. MR **28**:1199
- [Mi] Y. Miyaoka: *Themes and variations on inequalities of Chern classes*, Sugaku 41 (1989), 193–207. MR **91j**:14003
- [Pa] P. Pacelli: *Uniform boundedness for rational points*, Duke Math. J. 88 (1997), 77–102. MR **98b**:14020
- [P] A. N. Parshin: *Algebraic curves over function fields*, Izv. Akad. Nauk. SSSR Ser. Mat. 32 (1968), 1191–1219. MR **41**:1740
- [S] L. Szpiro: *Exposé No. 3 in Séminaire sur les pinceaux des courbes de genre au moins deux*, Astérisque 86 (1981). MR **83c**:14020

DEPARTMENT OF MATHEMATICS, UNIVERSITÀ DEGLI STUDI ROMA III, LARGO SAN LEONARDO MURIALDO 1, 00146 ROME, ITALY

E-mail address: caporaso@mat.uniroma3.it

IRREDUCIBILITY OF EQUISINGULAR FAMILIES OF CURVES

THOMAS KEILEN

ABSTRACT. In 1985 Joe Harris proved the long-standing claim of Severi that equisingular families of plane nodal curves are irreducible whenever they are nonempty. For families with more complicated singularities this is no longer true. Given a divisor D on a smooth projective surface Σ it thus makes sense to look for conditions which ensure that the family $V_{|D|}^{irr}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ of irreducible curves in the linear system $|D|_l$ with precisely r singular points of types $\mathcal{S}_1, \dots, \mathcal{S}_r$ is irreducible. Considering different surfaces, including general surfaces in $\mathbb{P}_{\mathbb{C}}^3$ and products of curves, we produce a sufficient condition of the type

$$\sum_{i=1}^r \deg(X(\mathcal{S}_i))^2 < \gamma \cdot (D - K_{\Sigma})^2,$$

where γ is some constant and $X(\mathcal{S}_i)$ some zero-dimensional scheme associated to the singularity type. Our results carry the same asymptotics as the best known results in this direction in the plane case, even though the coefficient is worse. For most of the surfaces considered these are the only known results in that direction.

1. INTRODUCTION

Equisingular families of curves have been studied quite intensively since the last century. If we fix a linear system $|D|_l$ on a smooth projective surface Σ and singularity types $\mathcal{S}_1, \dots, \mathcal{S}_r$, then we denote by $V^{irr} = V_{|D|}^{irr}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ the variety of irreducible curves in $|D|_l$ with precisely r singular points of the given types. The main questions are whether the equisingular family V^{irr} is nonempty, smooth of the expected dimension, and irreducible. For results in the plane case we refer to [GLS98c], [GLS00], and results on the first and the second question on other surfaces may be found in [GLS97], [GLS98a], [CC99], [Fla01], [Che01], [KT02]. In this paper, for the first time, the question of the irreducibility of V^{irr} for a wider range of surfaces is studied. As already families of cuspidal curves in the plane (cf. [Zar35]) or nodal curves on surfaces in $\mathbb{P}_{\mathbb{C}}^3$ (cf. [CC99]) show, in general we cannot expect a complete answer as for families of plane nodal curves, saying that the family is irreducible whenever it is nonempty. All we may hope for are numerical conditions depending on invariants of the singularity types, the surface and the linear system, which ensure the irreducibility of V^{irr} .

Received by the editors August 10, 2001 and, in revised form, February 5, 2002.

2000 *Mathematics Subject Classification*. Primary 14H10, 14H15, 14H20; Secondary 14J26, 14J27, 14J28, 14J70.

Key words and phrases. Algebraic geometry, singularity theory.

The author was partially supported by the DFG-Schwerpunkt "Globale Methoden in der komplexen Geometrie". The author would like to thank the referee for pointing out Example 2.5.

The main condition which we get (cf. Section 2) looks like

(1.1)
$$\sum_{i=1}^r \deg(X(\mathcal{S}_i))^2 < \gamma \cdot (D - K_\Sigma)^2,$$

where γ is some constant. Applying the estimates (1.6) for $\deg(X(\mathcal{S}_i))$ from Subsection 1.3 we could replace (1.1) by

(1.2)
$$\sum_{i=1}^r \tau(\mathcal{S}_i)^2 < \frac{\gamma}{9} \cdot (D - K_\Sigma)^2$$

in the case of analytical types, and in the topological case by

(1.3)
$$\sum_{i=1}^r \left(\mu(\mathcal{S}_i) + \frac{4}{3}\right)^2 < \frac{4 \cdot \gamma}{9} \cdot (D - K_\Sigma)^2.$$

In this section we introduce the basic concepts and notation used throughout the paper, and we state several important known facts. Section 2 contains the main results and their proofs, omitting the technical details. These are presented in Section 3 and Section 4.

1.1. General Assumptions and Notation. Throughout this article Σ will denote a smooth projective surface over \mathbb{C} ; \mathbb{N} denotes the set of nonnegative integers.

We will denote by $\text{Div}(\Sigma)$ the group of divisors on Σ and by K_Σ its canonical divisor. If D is any divisor on Σ , $\mathcal{O}_\Sigma(D)$ shall be the corresponding invertible sheaf, and we will sometimes write $H^\nu(X, D)$ instead of $H^\nu(X, \mathcal{O}_X(D))$. A *curve* $C \subset \Sigma$ will be an effective (nonzero) divisor, that is, a one-dimensional locally principal scheme, not necessarily reduced; however, an *irreducible curve* shall be reduced by definition. $|D|_l$ denotes the system of curves linearly equivalent to D . We will use the notation $\text{Pic}(\Sigma)$ for the *Picard group* of Σ , that is, $\text{Div}(\Sigma)$ modulo linear equivalence (denoted by \sim_l), and $\text{NS}(\Sigma)$ for the *Néron-Severi group*, that is, $\text{Div}(\Sigma)$ modulo algebraic equivalence (denoted by \sim_a). Given a reduced curve $C \subset \Sigma$, we will write $g(C)$ for its *geometric genus*.

Given any closed subscheme X of a scheme Y , we denote by $\mathcal{I}_X = \mathcal{I}_{X/Y}$ the *ideal sheaf* of X in \mathcal{O}_Y . If X is zero-dimensional, we denote by $\#X$ the number of points in its *support* $\text{supp}(X)$ and by $\deg(X) = \sum_{z \in Y} \dim_{\mathbb{C}}(\mathcal{O}_{Y,z}/\mathcal{I}_{X/Y,z})$ its *degree*.

If $X \subset \Sigma$ is a zero-dimensional scheme on Σ and $D \in \text{Div}(\Sigma)$, we denote by $|\mathcal{I}_{X/\Sigma}(D)|_l$ the linear system of curves C in $|D|_l$ with $X \subset C$.

If $L \subset \Sigma$ is any reduced curve and $X \subset \Sigma$ a zero-dimensional scheme, we define the *residue scheme* $X : L \subset \Sigma$ of X by the ideal sheaf $\mathcal{I}_{X:L/\Sigma} = \mathcal{I}_{X/\Sigma} : \mathcal{I}_{L/\Sigma}$ with stalks

$$\mathcal{I}_{X:L/\Sigma,z} = \mathcal{I}_{X/\Sigma,z} : \mathcal{I}_{L/\Sigma,z},$$

where “:” denotes the ideal quotient. This leads to the definition of the *trace scheme* $X \cap L \subset L$ of X via the ideal sheaf $\mathcal{I}_{X \cap L/L}$ given by the exact sequence

$$0 \longrightarrow \mathcal{I}_{X:L/\Sigma}(-L) \xrightarrow{\cdot L} \mathcal{I}_{X/\Sigma} \longrightarrow \mathcal{I}_{X \cap L/L} \longrightarrow 0.$$

1.2. Singularity Types. The germ $(C, z) \subset (\Sigma, z)$ of a reduced curve $C \subset \Sigma$ at a point $z \in \Sigma$ is called a *plane curve singularity*, and two plane curve singularities (C, z) and (C', z') are said to be *topologically* (respectively, *analytically equivalent*) if there is a homeomorphism (respectively, an analytical isomorphism) $\Phi : (\Sigma, z) \rightarrow (\Sigma, z')$ such that $\Phi(C) = C'$. We call an equivalence class with respect to these equivalence relations a *topological* (respectively, *analytical*) *singularity type*. The following are known to be invariants of the topological type \mathcal{S} of the plane curve singularity (C, z) : $r(\mathcal{S}) = r(C, z)$, the number of branches of (C, z) ; $\tau^{es}(\mathcal{S}) = \tau^{es}(C, z)$, the codimension of the μ -constant stratum in the semiuniversal deformation of (C, z) ; $\delta(\mathcal{S}) = \delta(C, z) = \dim_{\mathbb{C}}(\nu_* \mathcal{O}_{\tilde{C}, z} / \mathcal{O}_{C, z})$, the *delta invariant* of \mathcal{S} , where $\nu : (\tilde{C}, z) \rightarrow (C, z)$ is a normalisation of (C, z) ; and $\mu(\mathcal{S}) = \mu(C, z) = \dim_{\mathbb{C}} \mathcal{O}_{\Sigma, z} / (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$, the *Milnor number* of \mathcal{S} , where $f \in \mathcal{O}_{\Sigma, z}$ denotes a local equation of (C, z) with respect to the local coordinates x and y . For the analytical type \mathcal{S} of (C, z) we have as additional invariant, the *Tyurina number* of \mathcal{S} , defined as $\tau(\mathcal{S}) = \tau(C, z) = \dim_{\mathbb{C}} \mathcal{O}_{\Sigma, z} / (f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$. We recall the relation $2\delta(\mathcal{S}) = \mu(\mathcal{S}) + r(\mathcal{S}) - 1$ (cf. [Mil68, Chapter 10]). Furthermore, since the δ -constant stratum of the semiuniversal deformation of (C, z) contains the μ -constant stratum and since its codimension is just $\delta(\mathcal{S})$, we have $\delta(\mathcal{S}) \leq \tau^{es}(\mathcal{S})$ (see also [DH88]); hence

$$(1.4) \quad \mu(\mathcal{S}) \leq 2\delta(\mathcal{S}) \leq 2\tau^{es}(\mathcal{S}).$$

1.3. Singularity Schemes. For a reduced curve $C \subset \Sigma$ we recall the definition of the zero-dimensional schemes $X^{es}(C) \subseteq X^s(C)$ and $X^{ea}(C) \subseteq X^a(C)$ from [GLS00]. They are defined by the ideal sheaves $\mathcal{J}_{X^{es}(C)/\Sigma}$, $\mathcal{J}_{X^s(C)/\Sigma}$, $\mathcal{J}_{X^{ea}(C)/\Sigma}$, and $\mathcal{J}_{X^a(C)/\Sigma}$, respectively, given by the following stalks:

- $\mathcal{J}_{X^{es}(C)/\Sigma, z} = I^{es}(C, z) = \{g \in \mathcal{O}_{\Sigma, z} \mid f + \varepsilon g \text{ is equisingular over } \mathbb{C}[\varepsilon]/(\varepsilon^2)\}$, where $f \in \mathcal{O}_{\Sigma, z}$ is a local equation of C at z . $I^{es}(C, z)$ is called the *equisingularity ideal* of (C, z) .
- $\mathcal{J}_{X^s(C)/\Sigma, z} = \left\{g \in \mathcal{O}_{\Sigma, z} \mid g \text{ goes through the cluster } \mathcal{Cl}(C, T^*(C, z))\right\}$, where $T^*(C, z)$ denotes the essential subtree of the complete embedded resolution tree of (C, z) .
- $\mathcal{J}_{X^{ea}(C)/\Sigma, z} = I^{ea}(C, z) = (f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \subseteq \mathcal{O}_{\Sigma, z}$, where x, y denote local coordinates of Σ at z and $f \in \mathcal{O}_{\Sigma, z}$ is a local equation of C . $I^{ea}(C, z)$ is called the *Tyurina ideal* of (C, z) .
- $\mathcal{J}_{X^a(C)/\Sigma, z} = I^a(C, z) \subseteq \mathcal{O}_{\Sigma, z}$, where we refer for the somewhat lengthy definition of $I^a(C, z)$ to [GLS00, Section 1.3].

We call $X^{es}(C)$ the *equisingularity scheme* of C and $X^s(C)$ its *singularity scheme*. Analogously we call $X^{ea}(C)$ the *equianalytical singularity scheme* of C and $X^a(C)$ its *analytical singularity scheme*.

Throughout this article we will frequently treat topological and analytical singularities at the same time. Whenever we do so, we will write $X^*(C)$ for $X^{es}(C)$, respectively for $X^{ea}(C)$, and similarly $X(C)$ for $X^s(C)$, respectively for $X^a(C)$.

In [Los98], Propositions 2.19 and 2.20 and in Remarks 2.40 (see also [GLS00]) and 2.41, it is shown that, fixing a point $z \in \Sigma$ and a topological (respectively, analytical) type \mathcal{S} , the singularity schemes (respectively, analytical singularity schemes) having the same topological (respectively, analytical) type are parametrised by an irreducible Hilbert scheme, which we are going to denote by $\text{Hilb}_z(\mathcal{S})$. This then leads to an irreducible family

$$(1.5) \quad \text{Hilb}(\mathcal{S}) = \coprod_{z \in \Sigma} \text{Hilb}_z(\mathcal{S}).$$

In particular, equisingular (respectively, equianalytical) singularities have singularity schemes (respectively, analytical singularity schemes) of the same degree (see also [GLS98c] or [Los98], Lemma 2.8). The same is of course true regarding the equisingularity scheme (respectively, the equianalytical singularity scheme). If $C \subset \Sigma$ is a reduced curve such that z is a singular point of topological (respectively, analytical) type \mathcal{S} , we may therefore define $\deg(X(\mathcal{S})) = \deg(X(C), z)$ and $\deg(X^*(\mathcal{S})) = \deg(X^*(C), z)$. We note that, with this notation, $\dim \text{Hilb}_z(\mathcal{S}) = \deg(X(\mathcal{S})) - \deg(X^*(\mathcal{S})) - 2$ for any $z \in \Sigma$, and thus

$$\dim \text{Hilb}(\mathcal{S}) = \deg(X(\mathcal{S})) - \deg(X^*(\mathcal{S})).$$

In applications it is convenient to replace the degree of an (analytical) singularity scheme by an upper bound in known invariants of the singularities. From [Los98], p. 28, p. 103, and Lemma 2.44, it follows that for a topological (respectively, analytical) singularity type \mathcal{S} one has

$$(1.6) \quad \deg(X^a(\mathcal{S})) \leq 3\tau(\mathcal{S}) \quad \text{and} \quad \deg(X^s(\mathcal{S})) \leq \frac{3}{2}\mu(\mathcal{S}) + 2.$$

1.4. Equisingular Families. Given a divisor $D \in \text{Div}(\Sigma)$ and topological or analytical singularity types $\mathcal{S}_1, \dots, \mathcal{S}_r$, we denote by $V = V_{|D|}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ the locally closed subspace of $|D|_l$ of reduced curves in the linear system $|D|_l$ having precisely r singular points of types $\mathcal{S}_1, \dots, \mathcal{S}_r$. By $V^{reg} = V_{|D|}^{reg}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ we denote the open (cf. the proof of Theorem 3.1) subset

$$V^{reg} = \{C \in V \mid h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) = 0\} \subseteq V.$$

Similarly, we use the notation $V^{irr} = V_{|D|}^{irr}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ to denote the open subset of irreducible curves in the space V , and we set $V^{irr, reg} = V_{|D|}^{irr, reg}(\mathcal{S}_1, \dots, \mathcal{S}_r) = V^{irr} \cap V^{reg}$, which is open in V^{reg} and in V . If a type \mathcal{S} occurs $k > 1$ times, we rather write $k\mathcal{S}$ than $\mathcal{S}, \dots, \mathcal{S}$. We call these families of curves *equisingular families of curves*.

We say that V is *T-smooth* at $C \in V$ if the germ (V, C) is smooth of the (expected) dimension $\dim |D|_l - \deg(X^*(C))$.

By [Los98], Proposition 2.1 (see also [GK89], [GL96], [GLS00]) T-smoothness of V at C follows from the vanishing of $H^1(\Sigma, \mathcal{J}_{X^*(C)/\Sigma}(C))$, since the tangent space of V at C may be identified with $H^0(\Sigma, \mathcal{J}_{X^*(C)/\Sigma}(C))/H^0(\Sigma, \mathcal{O}_\Sigma)$.

¹ V^{reg} should not be confused with $\{C \in V \mid h^1(\Sigma, \mathcal{J}_{X^*(C)/\Sigma}(D)) = 0\}$, which is the part of V where V is smooth of the expected dimension. Curves in the latter subscheme are often called *regular* (c.f. [CC99]). See also Example 2.5.

1.5. Fibrations. Let $D \in \text{Div}(\Sigma)$ be a divisor, $\mathcal{S}_1, \dots, \mathcal{S}_r$ distinct topological or analytical singularity types, and $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$. We denote by \tilde{B} the irreducible parameter space

$$\tilde{B} = \tilde{B}(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r) = \prod_{i=1}^r \text{Sym}^{k_i}(\text{Hilb}(\mathcal{S}_i)),$$

and by $B = B(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r)$ the nonempty, open, irreducible and dense subspace

$$B = \left\{ ([X_{1,1}, \dots, X_{1,k_1}], \dots, [X_{r,1}, \dots, X_{r,k_r}]) \in \tilde{B} \mid \begin{aligned} &\text{supp}(X_{i,j}) \cap \text{supp}(X_{s,t}) = \emptyset \\ &\forall 1 \leq i, s \leq r, 1 \leq j \leq k_i, 1 \leq t \leq k_s \end{aligned} \right\}.$$

Note that $\dim(B)$ does not depend on Σ ; more precisely, with the notation of Subsection 1.3 we have

$$\dim(B) = \sum_{i=1}^r k_i \cdot (\deg(X(\mathcal{S}_i)) - \deg(X^*(\mathcal{S}_i))).$$

Let us set $n = \sum_{i=1}^r k_i \deg(X(\mathcal{S}_i))$. We then define an injective morphism

$$\begin{aligned} \psi = \psi(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r) : B(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r) &\longrightarrow \text{Hilb}_{\Sigma}^n, \\ ([X_{1,1}, \dots, X_{1,k_1}], \dots, [X_{r,1}, \dots, X_{r,k_r}]) &\longmapsto \bigcup_{i=1}^r \bigcup_{j=1}^{k_i} X_{i,j}, \end{aligned}$$

where Hilb_{Σ}^n denotes the smooth connected Hilbert scheme of zero-dimensional schemes of degree n on Σ (cf. [Los98], Section 1.3.1).

We denote by $\Psi = \Psi_D(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r)$ the fibration of $V_{|D|}(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r)$ induced by $B(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r)$; in other words, the morphism Ψ is given by

$$\begin{aligned} \Psi : V_{|D|}(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r) &\longrightarrow B(k_1 \mathcal{S}_1, \dots, k_r \mathcal{S}_r), \\ C &\longmapsto ([X_{1,1}, \dots, X_{1,k_1}], \dots, [X_{r,1}, \dots, X_{r,k_r}]), \end{aligned}$$

where $\text{Sing}(C) = \{z_{i,j} \mid i = 1, \dots, r, j = 1, \dots, k_i\}$, $X_{i,j} = X(C, z_{i,j})$ and $(C, z_{i,j}) \cong \mathcal{S}_i$ for all $i = 1, \dots, r, j = 1, \dots, k_i$.

With the notation of Subsection 1.4 note that for $C \in V$ the fibre $\Psi^{-1}(\Psi(C))$ is the open dense subset of the linear system $|\mathcal{J}_{X(C)/\Sigma}(D)|_t$ consisting of the curves C' with $X(C') = X(C)$. In particular, the fibres of Ψ restricted to V^{reg} are irreducible, and since for $C \in V^{reg}$ the cohomology group $H^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D))$ vanishes, they are equidimensional of dimension

$$h^0(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) - 1 = h^0(\Sigma, \mathcal{O}_{\Sigma}(D)) - \sum_{i=1}^r k_i \deg(X(\mathcal{S}_i)) - 1.$$

2. THE MAIN RESULTS

In this section we give sufficient conditions for the irreducibility of equisingular families of curves on certain surfaces with Picard number one, including the projective plane, general surfaces in $\mathbb{P}_{\mathbb{C}}^{3^3}$ and general K3-surfaces, on products of curves, and on a subclass of geometrically ruled surfaces.

2.1. Surfaces with Picard Number One.

Theorem 2.1. *Let Σ be a surface such that*

- (i) $\text{NS}(\Sigma) = L \cdot \mathbb{Z}$ with L ample, and
- (ii) $h^1(\Sigma, C) = 0$, whenever C is effective.

Let $D \in \text{Div}(\Sigma)$, let S_1, \dots, S_r be pairwise distinct topological or analytical singularity types and let $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$.

Suppose that

$$(2.1) \quad D - K_\Sigma \text{ is big and nef,}$$

$$(2.2) \quad D + K_\Sigma \text{ is nef,}$$

$$(2.3) \quad \sum_{i=1}^r k_i \deg(X(S_i)) < \beta \cdot (D - K_\Sigma)^2 \quad \text{for some } 0 < \beta \leq \frac{1}{4}, \text{ and}$$

$$(2.4) \quad \sum_{i=1}^r k_i \deg(X(S_i))^2 < \gamma \cdot (D - K_\Sigma)^2, \text{ where}$$

$$\gamma = \frac{(1 + \sqrt{1 - 4\beta})^2 \cdot L^2}{4 \cdot \chi(\mathcal{O}_\Sigma) + \max\{0, 2 \cdot K_\Sigma \cdot L\} + 6 \cdot L^2}.$$

Then either $V_{|D|}^{\text{irr}}(k_1 S_1, \dots, k_r S_r)$ is empty or it is irreducible of the expected dimension. \square

Remark 2.2. If we set

$$\gamma = \frac{36\alpha}{(3\alpha + 4)^2} \quad \text{with} \quad \alpha = \frac{4 \cdot \chi(\mathcal{O}_\Sigma) + \max\{0, 2 \cdot K_\Sigma \cdot L\} + 6 \cdot L^2}{L^2},$$

then a simple calculation shows that (2.3) becomes redundant. For this we have to take into account that $\deg(X(S)) \geq 3$ for any singularity type S . The claim then follows with $\beta = \frac{1}{3} \cdot \gamma \leq \frac{1}{4}$. \square

We now apply the result in several special cases.

Corollary 2.3. *Let $d \geq 3$, let $L \subset \mathbb{P}_{\mathbb{C}}^{3^2}$ be a line, and let S_1, \dots, S_r be topological or analytical singularity types.*

Suppose that

$$\sum_{i=1}^r \deg(X(S_i))^2 < \frac{90}{289} \cdot (d + 3)^2.$$

Then either $V_{|dL|}^{\text{irr}}(S_1, \dots, S_r)$ is empty or it is irreducible and T -smooth. \square

Many authors were concerned with the question in the case of plane curves with nodes and cusps or with nodes and one more complicated singularity or simply with ordinary multiple points; cf. e.g. [Sev21], [AC83], [Har85], [Kan89a], [Kan89b], [Ran89], [Shu91b], [Shu91a], [Bar93], [Shu94], [Shu96b], [Shu96a], [Wal96], [Los98], [GLS98a], [GLS98b], [Bru99], [GLS00]. Using techniques particularly designed for these cases they get of course better results than we may expect to.

The best general results in this case can be found in [GLS00] (see also [Los98], Corollary 6.1). Given a plane curve of degree d , omitting nodes and cusps, they get

$$\sum_{i=1}^r (\tau^*(S_i) + 2)^2 \leq \frac{9}{10} \cdot d^2$$

as the main irreducibility condition, where $\tau^*(\mathcal{S}_i) = \tau(\mathcal{S}_i)$ in the analytical case (respectively, $\tau^*(\mathcal{S}_i) = \tau^{es}(\mathcal{S}_i)$ in the topological case). By Subsection 1.2 we know that $\mu(\mathcal{S}_i) \leq 2 \cdot \tau^{es}(\mathcal{S}_i)$. Thus, in view of (1.2), (1.3), (1.4) and of Theorem 2.1 we get the sufficient condition

$$\sum_{i=1}^r \left(\tau^*(\mathcal{S}_i) + \frac{2}{3} \right)^2 < \frac{10}{289} \cdot (d+3)^2,$$

which has the same asymptotics. However, the coefficients differ by a factor of about 26.

A smooth complete intersection surface with Picard number one satisfies the assumptions of Theorem 2.1. Thus by the Noether Theorem of the result applies in particular to general surfaces in $\mathbb{P}_{\mathbb{C}}^3$.

Corollary 2.4. *Let $\Sigma \subset \mathbb{P}_{\mathbb{C}}^3$ be a smooth hypersurface of degree $n \geq 4$, let $H \subset \Sigma$ be a hyperplane section, and suppose that the Picard number of Σ is one. Let $d > n-4$ and let $\mathcal{S}_1, \dots, \mathcal{S}_r$ be topological or analytical singularity types.*

Suppose that

$$\sum_{i=1}^r \deg(X(\mathcal{S}_i))^2 < \frac{6 \cdot (n^3 - 3n^2 + 8n - 6) \cdot n^2}{(n^3 - 3n^2 + 10n - 6)^2} \cdot (d+4-n)^2.$$

Then either $V_{|dH|}^{irr}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ is empty or it is irreducible of the expected dimension. \square

We would like to thank the referee for pointing out the following example of reducible families $V_{|H|}^{irr}(3A_1)$ of nodal curves on surfaces in $\mathbb{P}_{\mathbb{C}}^3$.

Example 2.5. If $\Sigma \subset \mathbb{P}_{\mathbb{C}}^3$ is a general surface of degree $n \geq 4$, then there is a finite number $N > 1$ of 3-tangent planes to Σ . However, every 3-tangent plane cuts out an irreducible 3-nodal curve on Σ , and since the Picard group is generated by a hyperplane section H , every 3-nodal curve is of this form. Therefore, $V_{|H|}^{irr}(3A_1)$ consists of N distinct points. It is thus reducible, but smooth of the expected dimension

$$\dim(V_{|H|}^{irr}(3A_1)) = \dim |H|_l - 3 = 0.$$

Note that in this situation for $C \in V_{|H|}^{irr}(3A_1)$ and $z \in \text{Sing}(C)$ we have $\mathcal{J}_{X(C)/\Sigma, z} = \mathfrak{m}_{\Sigma, z}^2$ and thus

$$h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(H)) = 6 > 0.$$

Therefore, $V_{|H|}^{irr, reg}(3A_1) = \emptyset$. The parameter space B is just $\text{Sym}^3(\Sigma)$.

A general K3-surface has Picard number one and in this situation, by the Kodaira Vanishing Theorem, Σ also satisfies the assumption (ii) in Theorem 2.1.

Corollary 2.6. *Let Σ be a smooth K3-surface with $\text{NS}(\Sigma) = L \cdot \mathbb{Z}$ with L ample and set $n = L^2$. Let $d > 0$, $D \sim_a dL$ and let $\mathcal{S}_1, \dots, \mathcal{S}_r$ be topological or analytical singularity types.*

Suppose that

$$\sum_{i=1}^r \deg(X(\mathcal{S}_i))^2 < \frac{54n^2 + 72n}{(11n + 12)^2} \cdot d^2 \cdot n.$$

Then either $V_{|D|}^{irr}(\mathcal{S}_1, \dots, \mathcal{S}_r)$ is empty or it is irreducible of the expected dimension. \square

2.2. Products of Curves. If $\Sigma = C_1 \times C_2$ is the product of two smooth projective curves, then for a general choice of C_1 and C_2 the Néron–Severi group will be generated by two fibres of the canonical projections, by abuse of notation also denoted by C_1 and C_2 . If both curves are elliptic, then “general” just means that the two curves are non-isogenous.

Theorem 2.7. *Let C_1 and C_2 be two smooth projective curves of genera g_1 and g_2 , respectively, with $g_1 \geq g_2 \geq 0$, such that for $\Sigma = C_1 \times C_2$ the Néron–Severi group is $\text{NS}(\Sigma) = C_1\mathbb{Z} \oplus C_2\mathbb{Z}$.*

Let $D \in \text{Div}(\Sigma)$ such that $D \sim_a aC_1 + bC_2$ with $a > \max\{2g_2 - 2, 2 - 2g_2\}$ and $b > \max\{2g_1 - 2, 2 - 2g_1\}$, let $\mathcal{S}_1, \dots, \mathcal{S}_r$ be pairwise distinct topological or analytical singularity types and $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$.

Suppose that

$$(2.5) \quad \sum_{i=1}^r k_i \deg(X(\mathcal{S}_i))^2 < \gamma \cdot (D - K_\Sigma)^2,$$

where γ may be taken from the following table with $\alpha = \frac{a-2g_2+2}{b-2g_1+2} > 0$.

g_1	g_2	γ
0	0	$\frac{1}{24}$
1	0	$\frac{1}{\max\{32, 2\alpha\}}$
≥ 2	0	$\frac{1}{\max\{24+16g_1, 4g_1\alpha\}}$
1	1	$\frac{1}{\max\{32, 2\alpha, \frac{2}{\alpha}\}}$
≥ 2	≥ 1	$\frac{1}{\max\{24+16g_1+16g_2, 4g_1\alpha, \frac{4g_2}{\alpha}\}}$

Then either $V_{|D|}^{irr}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ is empty or it is irreducible of the expected dimension. \square

Only in the case $\Sigma \cong \mathbb{P}_{\mathbb{C}}^{3^1} \times \mathbb{P}_{\mathbb{C}}^{3^1}$ we get a constant γ which does not depend on the chosen divisor D , while in the remaining cases the ratio of a and b is involved in γ . This means that an asymptotical behaviour can only be examined if the ratio remains unchanged.

2.3. Geometrically Ruled Surfaces. Let $\pi : \Sigma = \mathbb{P}(\mathcal{E}) \rightarrow C$ be a geometrically ruled surface with normalised bundle \mathcal{E} (in the sense of [Har77], V.2.8.1). The Néron–Severi group of Σ is $\text{NS}(\Sigma) = C_0\mathbb{Z} \oplus F\mathbb{Z}$ with intersection matrix $\begin{pmatrix} -e & 1 \\ 1 & 0 \end{pmatrix}$, where $F \cong \mathbb{P}_{\mathbb{C}}^{3^1}$ is a fibre of π , C_0 a section of π with $\mathcal{O}_{\Sigma}(C_0) \cong \mathcal{O}_{\mathbb{P}(\mathcal{E})}(1)$, $g = g(C)$ the genus of C , $e = \Lambda^2 \mathcal{E}$ and $e = -\deg(e) \geq -g$. For the canonical divisor we have $K_\Sigma \sim_a -2C_0 + (2g - 2 - e) \cdot F$.

Theorem 2.8. *Let $\pi : \Sigma \rightarrow C$ be a geometrically ruled surface with $e \leq 0$. Let $D = aC_0 + bF \in \text{Div}(\Sigma)$ with $a \geq 2$, $b > 2g - 2 + \frac{ae}{2}$, and if $g = 0$, then $b \geq 2$. Let $\mathcal{S}_1, \dots, \mathcal{S}_r$ be pairwise distinct topological or analytical singularity types and $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$.*

Suppose that

$$(2.6) \quad \sum_{i=1}^r k_i \deg (X(\mathcal{S}_i))^2 < \gamma \cdot (D - K_\Sigma)^2,$$

where γ may be taken from the following table with $\alpha = \frac{a+2}{b+2-2g-\frac{ae}{2}} > 0$.

g	e	γ
0	0	$\frac{1}{24}$
1	0	$\frac{1}{\max\{24, 2\alpha\}}$
1	-1	$\frac{1}{\max\left\{\min\left\{30+\frac{16}{\alpha}+4\alpha, 40+9\alpha\right\}, \frac{13}{2}\alpha\right\}}$
≥ 2	0	$\frac{1}{\max\{24+16g, 4g\alpha\}}$
≥ 2	< 0	$\frac{1}{\max\left\{\min\left\{24+16g-9e\alpha, 18+16g-9e\alpha-\frac{16}{e\alpha}\right\}, 4g\alpha-9e\alpha\right\}}$

Then either $V_{|D|}^{irr}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ is empty or it is irreducible of the expected dimension. \square

Once more, only in the case $g = 0$, i.e., when $\Sigma \cong \mathbb{P}_\mathbb{C}^{3^1} \times \mathbb{P}_\mathbb{C}^{3^1}$, we are in the lucky situation that the constant γ does not at all depend on the chosen divisor D , whereas in the case $g \geq 1$ the ratio of a and b is involved in γ . This means that an asymptotical behaviour can only be examined if the ratio remains unchanged.

If Σ is a product $C \times \mathbb{P}_\mathbb{C}^{3^1}$, the constant γ here is the same as in Section 2.2.

In [Ran89] and in [GLS98a] the case of nodal curves on the Hirzebruch surface \mathbb{F}_1 is treated, since this is just $\mathbb{P}_\mathbb{C}^{3^2}$ blown up at one point. \mathbb{F}_1 is an example of a geometrically ruled surface with invariant $e = 1 > 0$, a case which we so far cannot treat with our methods, due to the section with self-intersection -1 . However, it seems to be possible to extend the methods of [GLS98a] to the situation of arbitrary ruled surfaces with positive invariant e , at least if we restrict ourselves to singularities that are not too bad.

2.4. The Proofs. Our approach to the problem proceeds along the lines of an unpublished result of Greuel, Lossen and Shustin (cf. [GLS98b]), which is based on ideas of Chiantini and Ciliberto (cf. [CC99]). The basic ideas are in some respect similar to the approach used in [GLS00], replacing the ‘‘Castelnuovo-function’’ arguments by ‘‘Bogomolov instability’’.

We first show that the open subscheme $V_{|D|}^{irr,reg} = V_{|D|}^{irr,reg}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ of $V^{irr} = V_{|D|}^{irr}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$, and hence its closure $\overline{V^{irr,reg}}$ in V^{irr} , is always irreducible (cf. Theorem 3.1), and then we look for criteria which ensure that the complement of $\overline{V^{irr,reg}}$ in V^{irr} is empty (cf. Section 4). For the latter, we consider the restriction of the morphism $\Psi : V \rightarrow B$ (cf. Subsection 1.5) to an irreducible component V^* of V^{irr} not contained in $\overline{V^{irr,reg}}$. From the fact that the dimension of V^* is at least the expected dimension $\dim(V^{irr,reg})$, we deduce that the codimension of $B^* = \Psi(V^*)$ in B is at most $h^1(\Sigma, \mathcal{I}_{X(C)/\Sigma}(D))$, where $C \in V^*$ (cf. Lemma 4.7). It thus suffices to find conditions that contradict this inequality, that is, we have to get our hands on $\text{codim}_B(B^*)$. However, on the surfaces that

we consider, the non-vanishing of $h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D))$ means in some sense that the zero-dimensional scheme $X(C)$ is in a special position. We may thus hope to realise large parts X_i^0 of $X(C)$ on curves Δ_i of “small degree” ($i = 1, \dots, m$), which would impose at least $\#X_i^0 - \dim |\Delta_i|_l$ conditions on $X(C)$, giving rise to a lower bound $\sum_{i=1}^m \#X_i^0 - \dim |\Delta_i|_l$ for $\text{codim}_B(B^*)$. The X_i^0 ’s and the Δ_i ’s are found in Lemma 4.1 with the aid of certain Bogomolov unstable rank-two bundles. It thus finally remains (cf. Lemmata 4.3, 4.4 and 4.6) to give conditions that imply

$$\sum_{i=1}^m \#X_i^0 - \dim |\Delta_i|_l > h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)).$$

These considerations lead to the following proofs.

Proof of Theorem 2.1. We may assume that V^{irr} is nonempty. By Theorem 3.1 it suffices to show that $V^{irr} = \overline{V^{irr, reg}}$.

Suppose the contrary, i.e., there is an irreducible curve $C_0 \in V^{irr} \setminus \overline{V^{irr, reg}}$; in particular, $h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)) > 0$ for $X_0 = X(C_0)$. Since

$$\deg(X_0) = \sum_{i=1}^r k_i \deg(X(\mathcal{S}_i)),$$

and $\sum_{z \in \Sigma} (\deg(X_{0,z}))^2 = \sum_{i=1}^r k_i \deg(X(\mathcal{S}_i))^2$, the assumptions (0)–(3) of Lemma 4.1 and (4) of Lemma 4.3 are fulfilled. Thus Lemma 4.3 implies that C_0 satisfies condition (4.19) in Lemma 4.7, which it cannot satisfy by the same lemma. Thus we have derived a contradiction. \square

Proof of Theorem 2.7. The assumptions on a and b ensure that $D - K_\Sigma$ is big and nef and that $D + K_\Sigma$ is nef. Thus, once we know that (2.5) implies condition (3) in Lemma 4.1 we can do the same proof as in Theorem 2.1, just replacing Lemma 4.3 by Lemma 4.4.

For condition (3) we note that

$$\sum_{i=1}^r k_i \deg(X(\mathcal{S}_i)) \leq \sum_{i=1}^r k_i \cdot \left(\deg(X(\mathcal{S}_i)) \right)^2 \leq \frac{1}{24} \cdot (D - K_\Sigma)^2 < \frac{1}{4} \cdot (D - K_\Sigma)^2.$$

\square

Proof of Theorem 2.8. The proof is identical to that of Theorem 2.7, just replacing Lemma 4.4 by Lemma 4.6. \square

2.5. Some Remarks. What are the obstructions to our approach?

First, the Bogomolov instability does not give much information about the curves Δ_i apart from their existence and the fact that they are in some sense “small” compared with the divisor D . We are thus bound to the study of surfaces where we have a good knowledge of the dimension of arbitrary complete linear systems. Second, in order to derive the above inequality, many nasty calculations are necessary which strongly depend on the particular structure of the Néron–Severi group of the surface, that is, we are restricted to surfaces where the Néron–Severi group is not too large and the intersection pairing is not too hard (cf. Lemmata 4.3, 4.4 and 4.6). Finally, in order to ensure the Bogomolov instability of the vector bundle considered throughout the proof of Lemma 4.1 we heavily use the fact that the surface Σ does not contain any curve of negative self-intersection, which excludes e.g. general Hirzebruch surfaces.

If the number of irreducible curves of negative self-intersection is not too large, one might overcome this last obstacle with the technique used in [GLS98a]. That is, we would have to show that under certain additional conditions the singular points of the considered curves could be independently moved; in particular, they could be moved off the exceptional curves—more precisely, the subvariety of V^{irr} of curves whose singular locus does not lie on any exceptional curve is dense in V^{irr} . For this, one basically just needs criteria for the existence of “small” curves realising a zero-dimensional scheme slightly bigger than the equisingularity scheme (respectively, the equianalytical singularity scheme) of the members in V^{irr} . For example, in the case of curves with r nodes, that means the existence of curves passing through r arbitrary points and having multiplicity two in one of them.

In Section 3 we not only prove that $V^{irr,reg}$ is irreducible, but also that this indeed remains true if we drop the requirement that the curves should be irreducible, i.e., we show that V^{reg} is irreducible. However, unfortunately our approach does not give conditions for the emptiness of the complement of $\overline{V^{reg}}$, and thus we cannot say anything about the irreducibility of the variety of possibly reducible curves in $|D|_l$ with prescribed singularities. The reason for this is that in the proof of Lemma 4.1 we use the Bézout Theorem to estimate $D \cdot \Delta_i$.

3. $V^{irr,reg}$ IS IRREDUCIBLE

We now show that $V^{irr,reg}$ is always irreducible. We do this by showing that under $\Psi : V \rightarrow B$ every irreducible component of $V^{irr,reg}$ is smooth and maps dominant to the irreducible variety B with irreducible fibres.

Theorem 3.1. *Let $D \in \text{Div}(\Sigma)$, $\mathcal{S}_1, \dots, \mathcal{S}_r$ be pairwise distinct topological or analytical singularity types and $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$.*

If $V^{irr,reg}_{|D|}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ is nonempty, then it is a T -smooth, irreducible, open subset of $V^{irr}_{|D|}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ of dimension $\dim |D|_l - \sum_{i=1}^r k_i \deg(X^(\mathcal{S}_i))$.*

Proof. Since $V^{irr,reg}_{|D|}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$ is an open subset of $V^{reg}_{|D|}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r) = V^{reg}$, it suffices to show the claim for V^{reg} .

Let us consider the following maps from Subsection 1.5:

$$\Psi = \Psi_D(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r) : V = V_{|D|}(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r) \longrightarrow B(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r)$$

and

$$\psi = \psi(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r) : B(k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r) \longrightarrow \text{Hilb}_{\Sigma}^n.$$

Step 1: Every irreducible component V^* of V^{reg} is T -smooth of dimension $\dim |D|_l - \sum_{i=1}^r k_i \deg(X^*(\mathcal{S}_i))$. By [Los98], Proposition 2.1 (c2), V^* is T -smooth at any $C \in V^*$ of dimension $\dim |D|_l - \deg(X^*(C))$, since $h^1(\Sigma, \mathcal{J}_{X^*/\Sigma}(D)) = 0$. Note that $\deg(X^*(C)) = \sum_{i=1}^r k_i \deg(X^*(\mathcal{S}_i))$ only depends on $k_1\mathcal{S}_1, \dots, k_r\mathcal{S}_r$ (cf. Subsection 1.3).

Step 2: V^{reg} is open in V . Let $C \in V^{reg}$; then $h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) = 0$. Thus by semicontinuity there exists an open, dense neighbourhood U of $X(C)$ in Hilb_{Σ}^n such that $h^1(\Sigma, \mathcal{J}_{Y/\Sigma}(D)) = 0$ for all $Y \in U$. But then $\Psi^{-1}(\psi^{-1}(U)) \subseteq V^{reg}$ is an open neighbourhood of C in V , and hence V^{reg} is open in V .

Step 3: Ψ restricted to any irreducible component V^* of V^{reg} is dominant. Let V^* be an irreducible component of V^{reg} and let $C \in V^*$. Since $\Psi^{-1}(\Psi(C))$ is an

open, dense subset of $|\mathcal{J}_{X(C)/\Sigma}(D)|_l$ and since $h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) = 0$, we have $\dim \Psi^{-1}(\Psi(C)) = h^0(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) - 1 = \dim |D|_l - \deg(X(C))$. By Step 1 we know the dimension of V^* and by Subsection 1.5 we also know the dimension of B . Thus we conclude

$$\begin{aligned} \dim \Psi(V^*) &= \dim V^* - \dim \Psi^{-1}(\Psi(C)) \\ &= (\dim |D|_l - \deg X^*(C)) - (\dim |D|_l - \deg X(C)) \\ &= \deg(X(C)) - \deg(X^*(C)) = \dim B. \end{aligned}$$

Since B is irreducible, $\Psi(V^*)$ must be dense in B .
Step 4: V^{reg} is irreducible. Let V^* and V^{**} be two irreducible components of V^{reg} . Then $\Psi(V^*) \cap \Psi(V^{**}) \neq \emptyset$, and thus some fibre F of Ψ intersects both, V^* and V^{**} . However, the fibre is irreducible and by Step 1 both V^* and V^{**} are smooth. Thus F must be completely contained in V^* and V^{**} , which implies that $V^* = V^{**}$, since both are smooth of the same dimension. Thus V^{reg} is irreducible. \square

4. THE TECHNICAL DETAILS

The following lemma is the heart of the proof. Given a curve $C \in |D|_l$, whose (analytical) singularity scheme $X_0 = X(C)$ is *special with respect to D* in the sense that $h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)) > 0$, provides a “small” curve Δ_1 through a subscheme X_1^0 of X_0 , so that we can reduce the problem by replacing X_0 and D by $X_0 : \Delta_1$ and $D - \Delta_1$ respectively. We can of course proceed inductively as long as the new zero-dimensional scheme is again special with respect to the new divisor.
In order to find Δ_1 we choose a subscheme $X_1^0 \subseteq X_0$ that is minimal among those subschemes special with respect to D . By Grothendieck–Serre duality,

$$H^1(\Sigma, \mathcal{J}_{X_1^0/\Sigma}(D)) \cong \text{Ext}^1(\mathcal{J}_{X_1^0/\Sigma}(D - K_\Sigma), \mathcal{O}_\Sigma)$$

and a nontrivial element of the latter group gives rise to an extension

$$0 \rightarrow \mathcal{O}_\Sigma \rightarrow E_1 \rightarrow \mathcal{J}_{X_1^0/\Sigma}(D - K_\Sigma) \rightarrow 0.$$

We then show that the rank-two bundle E_1 is Bogomolov unstable and deduce the existence of a divisor Δ_1^0 such that

$$H^0(\Sigma, \mathcal{J}_{X_1^0/\Sigma}(D - K_\Sigma - \Delta_1^0)) \neq 0,$$

that is, we find a curve $\Delta_1 \in |\mathcal{J}_{X_1^0/\Sigma}(D - K_\Sigma - \Delta_1^0)|_l$.

Lemma 4.1. *Let Σ be a surface such that any curve $C \subset \Sigma$ is nef. (Assumption $(*)$)*

Let $D \in \text{Div}(\Sigma)$ and $X_0 \subset \Sigma$ a zero-dimensional scheme satisfying

- (0) $D - K_\Sigma$ is big and nef, and $D + K_\Sigma$ is nef,*
- (1) $\exists C_0 \in |D|_l$ irreducible : $X_0 \subset C_0$,*
- (2) $h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)) > 0$, and*
- (3) $\deg(X_0) < \beta \cdot (D - K_\Sigma)^2$ for some $0 < \beta \leq \frac{1}{4}$.*

Then there exist curves $\Delta_1, \dots, \Delta_m \subset \Sigma$ and zero-dimensional locally complete intersections $X_i^0 \subseteq X_{i-1} \cap \Delta_i$ for $i = 1, \dots, m$, where $X_i = X_{i-1} : \Delta_i$ for $i = 1, \dots, m$, such that

$$(a) \quad h^1\left(\Sigma, \mathcal{J}_{X_m/\Sigma}(D - \sum_{i=1}^m \Delta_i)\right) = 0,$$

and for $i = 1, \dots, m$

- (b) $h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) = 1,$
- (c) $D \cdot \Delta_i \geq \deg(X_{i-1} \cap \Delta_i) \geq \deg(X_i^0) \geq (D - K_\Sigma - \sum_{k=1}^i \Delta_k) \cdot \Delta_i \geq \Delta_i^2 \geq 0,$
- (d) $(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i)^2 > 0,$
- (e) $(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i) \cdot H > 0$ for all $H \in \text{Div}(\Sigma)$ ample, and
- (f) $D - K_\Sigma - \sum_{k=1}^i \Delta_k$ is big and nef.

Moreover, it follows that

$$(4.1) \quad 0 \leq \frac{1}{4}(D - K_\Sigma)^2 - \sum_{i=1}^m \deg(X_i^0) \leq \left(\frac{1}{2}(D - K_\Sigma) - \sum_{i=1}^m \Delta_i \right)^2.$$

Proof. We are going to find the schemes Δ_i and X_i^0 recursively. Let us therefore suppose that we have already found $\Delta_1, \dots, \Delta_{i-1}$ and X_1^0, \dots, X_{i-1}^0 satisfying (b)–(f), and suppose that still $h^1\left(\Sigma, \mathcal{J}_{X_{i-1}/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) > 0$.

We choose minimal $X_i^0 \subseteq X_{i-1}$ such that $h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) > 0$.

Step 1: $h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) = 1$, i.e., (b) is fulfilled.

Suppose it was strictly larger than one. By (0), respectively (f), and by the Kawamata–Viehweg Vanishing Theorem we have $h^1\left(\Sigma, \mathcal{O}_\Sigma(D - \sum_{k=1}^{i-1} \Delta_k)\right) = 0$.

Thus X_i^0 cannot be empty, that is, $\deg(X_i^0) \geq 1$ and we may choose a subscheme $Y \subset X_i^0$ of degree $\deg(Y) = \deg(X_i^0) - 1$. The inclusion $\mathcal{J}_{X_i^0} \hookrightarrow \mathcal{J}_Y$ implies $h^0\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) \leq h^0\left(\Sigma, \mathcal{J}_{Y/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right)$ and the structure sequences of Y and X_i^0 thus lead to

$$h^1\left(\Sigma, \mathcal{J}_{Y/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) \geq h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) - 1 > 0$$

contradicting the minimality of X_i^0 .

Step 2: $\deg(X_i^0) \leq \deg(X_0) - \sum_{k=1}^{i-1} \deg(X_{k-1} \cap \Delta_k)$.

The case $i = 1$ follows from the fact that $X_1^0 \subseteq X_0$, and for $i > 1$ the inclusion $X_i^0 \subseteq X_{i-1} = X_{i-2} : \Delta_{i-1}$ implies

$$\deg(X_i^0) \leq \deg(X_{i-2} : \Delta_{i-1}) = \deg(X_{i-2}) - \deg(X_{i-2} \cap \Delta_{i-1}).$$

It thus suffices to show that

$$\deg(X_{i-2}) - \deg(X_{i-2} \cap \Delta_{i-1}) = \deg(X_0) - \sum_{k=1}^{i-1} \deg(X_{k-1} \cap \Delta_k).$$

If $i = 2$, there is nothing to show. Otherwise $X_{i-2} = X_{i-3} : \Delta_{i-2}$ implies

$$\begin{aligned} & \deg(X_{i-2}) - \deg(X_{i-2} \cap \Delta_{i-1}) \\ &= \deg(X_{i-3} : \Delta_{i-2}) - \deg(X_{i-2} \cap \Delta_{i-1}) \\ &= \deg(X_{i-3}) - \deg(X_{i-3} \cap \Delta_{i-2}) - \deg(X_{i-2} \cap \Delta_{i-1}) \end{aligned}$$

and we are done by induction.

Step 3: There exists a “suitable” locally free rank-two vector bundle E_i .

By the Grothendieck–Serre duality we have

$$0 \neq H^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) \cong \text{Ext}^1\left(\mathcal{J}_{X_i^0/\Sigma}(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k), \mathcal{O}_\Sigma\right).$$

That is, there exists an extension

$$(4.2) \quad 0 \rightarrow \mathcal{O}_\Sigma \rightarrow E_i \rightarrow \mathcal{J}_{X_i^0/\Sigma}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k\right) \rightarrow 0.$$

The minimality of X_i^0 implies that E_i is locally free and hence that X_i^0 is a locally complete intersection (cf. [Laz97]),

$$(4.3) \quad c_1(E_i) = D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k \quad \text{and} \quad c_2(E_i) = \deg(X_i^0).$$

Step 4: E_i is Bogomolov unstable.

According to the Bogomolov Theorem we only have to show that $c_1(E_i)^2 > 4c_2(E_i)$ (cf. [Bog79] or [Laz97], Theorem 4.2). Since $(4\beta - 1) \cdot (D - K_\Sigma)^2 \leq 0$ by (3) and since $\Delta_k^2 \geq 0$ by Assumption (*), we deduce:

$$\begin{aligned} 4c_2(E_i) &= 4 \deg(X_i^0) \stackrel{\text{Step 2}}{\leq} 4 \deg(X_0) - 4 \sum_{k=1}^{i-1} \deg(X_{k-1} \cap \Delta_k) \\ &\stackrel{(3)/(c)}{<} 4\beta(D - K_\Sigma)^2 - 2 \sum_{k=1}^{i-1} \Delta_k \cdot \left(D - K_\Sigma - \sum_{j=1}^k \Delta_j\right) - 2 \sum_{k=1}^{i-1} \Delta_k^2 \\ &= \left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k\right)^2 + (4\beta - 1) \cdot (D - K_\Sigma)^2 - \sum_{k=1}^{i-1} \Delta_k^2 \\ &\leq \left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k\right)^2 = c_1(E_i)^2. \end{aligned}$$

Step 5: Find Δ_i .

Since E_i is Bogomolov unstable, there is a 0-dimensional scheme $Z_i \subset \Sigma$ and a $\Delta_i^0 \in \text{Div}(\Sigma)$ such that

$$(4.4) \quad 0 \rightarrow \mathcal{O}_\Sigma(\Delta_i^0) \rightarrow E_i \rightarrow \mathcal{J}_{Z_i/\Sigma}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k - \Delta_i^0\right) \rightarrow 0$$

is exact and such that

$$(d') \quad (2\Delta_i^0 - D + K_\Sigma + \sum_{k=1}^{i-1} \Delta_k)^2 \geq c_1(E_i)^2 - 4 \cdot c_2(E_i) > 0, \text{ and}$$

$$(e') \quad (2\Delta_i^0 - D + K_\Sigma + \sum_{k=1}^{i-1} \Delta_k) \cdot H > 0 \quad \text{for all } H \in \text{Div}(\Sigma) \text{ ample.}$$

Tensoring (4.4) with $\mathcal{O}_\Sigma(-\Delta_i^0)$ leads to the following exact sequence:

$$(4.5) \quad 0 \rightarrow \mathcal{O}_\Sigma \rightarrow E_i(-\Delta_i^0) \rightarrow \mathcal{J}_{Z_i/\Sigma}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k - 2\Delta_i^0\right) \rightarrow 0,$$

and we deduce that $h^0(\Sigma, E_i(-\Delta_i^0)) \neq 0$.

Now tensoring (4.2) with $\mathcal{O}_\Sigma(-\Delta_i^0)$ leads to

$$(4.6) \quad 0 \rightarrow \mathcal{O}_\Sigma(-\Delta_i^0) \rightarrow E_i(-\Delta_i^0) \rightarrow \mathcal{J}_{X_i^0/\Sigma}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k - \Delta_i^0\right) \rightarrow 0.$$

By (e'), and (0), respectively (f),

$$-\Delta_i^0.H < -\frac{1}{2}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k\right).H \leq 0$$

for an ample divisor H . Hence $-\Delta_i^0$ cannot be effective, that is, $H^0(\Sigma, -\Delta_i^0) = 0$. But the long exact cohomology sequence of (4.6) then implies

$$0 \neq H^0\left(\Sigma, E_i(-\Delta_i^0)\right) \hookrightarrow H^0\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}\left(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k - \Delta_i^0\right)\right).$$

In particular, we may choose $\Delta_i \in \left|\mathcal{J}_{X_i^0/\Sigma}(D - K_\Sigma - \sum_{k=1}^{i-1} \Delta_k - \Delta_i^0)\right|_l$.

Step 6: Δ_i satisfies (d)–(f).

We note that by the choice of Δ_i we have the following equivalences:

$$(4.7) \quad \Delta_i^0 \sim_l D - K_\Sigma - \sum_{k=1}^i \Delta_k,$$

$$(4.8) \quad \Delta_i^0 - \Delta_i \sim_l 2\Delta_i^0 - D + K_\Sigma + \sum_{k=1}^{i-1} \Delta_k \sim_l D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i.$$

Thus (d) and (e) is a reformulation of (d') and (e').

Moreover, since $(\Delta_i^0 - \Delta_i).H > 0$ for any ample H , we have $(\Delta_i^0 - \Delta_i).H \geq 0$ for any H in the closure of the ample cone; in particular,

$$(4.9) \quad \Delta_i^0.H \geq \Delta_i.H \geq 0 \quad \text{for all } H \text{ nef,}$$

since Δ_i is effective. Finally, since by assumption (*) any effective divisor is nef, we deduce that $\Delta_i^0.C \geq 0$ for any curve C , that is, Δ_i^0 is nef. In view of (4.7) for (f) it remains to show that $(\Delta_i^0)^2 > 0$. Once more taking into account that Δ_i is nef by (*) we have by (d'), (4.8), and (4.9),

$$(\Delta_i^0)^2 = (\Delta_i^0 - \Delta_i)^2 + (\Delta_i^0 - \Delta_i).\Delta_i + \Delta_i^0.\Delta_i > 0.$$

Step 7: Δ_i satisfies (c).

We would like to apply the Bézout Theorem to C_0 and Δ_i . Thus suppose that the irreducible curve C_0 is a component of Δ_i and let H be any ample divisor.

Applying (d) and the fact that $D + K_\Sigma$ is nef by (0), we derive the contradiction

$$0 \leq (\Delta_i - C_0).H < -\frac{1}{2} \cdot \left(D + K_\Sigma + \sum_{k=1}^{i-1} \Delta_k\right).H \leq -\frac{1}{2} \cdot (D + K_\Sigma).H \leq 0.$$

Since $X_{i-1} \subseteq X_0 \subset C_0$, the Bézout Theorem therefore implies

$$D.\Delta_i = C_0.\Delta_i \geq \deg(X_{i-1} \cap \Delta_i).$$

By definition $X_i^0 \subseteq X_{i-1}$ and $X_i^0 \subset \Delta_i$. Thus

$$\deg(X_{i-1} \cap \Delta_i) \geq \deg(X_i^0).$$

By assumption (*) the curve Δ_i is nef and thus (4.9) gives

$$\left(D - K_\Sigma - \sum_{k=1}^i \Delta_k\right).\Delta_i = \Delta_i^0.\Delta_i \geq \Delta_i^2 \geq 0.$$

Finally from (d') and by (4.3) it follows that

$$(\Delta_i^0 - \Delta_i)^2 \geq c_1(E_i)^2 - 4 \cdot c_2(E_i) = (\Delta_i^0 + \Delta_i)^2 - 4 \cdot \deg(X_i^0),$$

and thus $\deg(X_i^0) \geq \Delta_i^0 \cdot \Delta_i$.

Step 8: After a finite number m of steps, $h^1(\Sigma, \mathcal{J}_{X_m/\Sigma}(D - \sum_{i=1}^m \Delta_i)) = 0$.

As we have mentioned in Step 1, $\deg(X_i^0) > 0$. This ensures that

$$\deg(X_i) = \deg(X_{i-1}) - \deg(X_{i-1} \cap \Delta_i) \leq \deg(X_{i-1}) - \deg(X_i^0) < \deg(X_{i-1}),$$

i.e., the degree of X_i strictly decreases each time. Thus the procedure must stop after a finite number m of steps

Step 9: It remains to show (4.1).

By assumption (*) the curves Δ_i are nef; in particular, $\Delta_i \cdot \Delta_j \geq 0$ for all i, j . Thus (c) implies

$$\begin{aligned} \sum_{i=1}^m \deg(X_i^0) &\geq \sum_{i=1}^m (D - K_\Sigma - \sum_{k=1}^i \Delta_k) \cdot \Delta_i \\ &= (D - K_\Sigma) \cdot \sum_{i=1}^m \Delta_i - \frac{1}{2} \left(\left(\sum_{i=1}^m \Delta_i \right)^2 + \sum_{i=1}^m \Delta_i^2 \right) \\ &\geq (D - K_\Sigma) \cdot \sum_{i=1}^m \Delta_i - \left(\sum_{i=1}^m \Delta_i \right)^2. \end{aligned}$$

But then, taking condition (3) into account,

$$\begin{aligned} 0 \leq \frac{1}{4}(D - K_\Sigma)^2 - \deg(X_0) &\leq \frac{1}{4}(D - K_\Sigma)^2 - \sum_{i=1}^m \deg(X_i^0) \\ &\leq \frac{1}{4}(D - K_\Sigma)^2 - (D - K_\Sigma) \cdot \sum_{i=1}^m \Delta_i + \left(\sum_{i=1}^m \Delta_i \right)^2 \\ &= \left(\frac{1}{2}(D - K_\Sigma) - \sum_{i=1}^m \Delta_i \right)^2. \end{aligned}$$

□

It is our overall aim to compare the dimension of a cohomology group of the form $H^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D))$ with some invariants of the X_i^0 and Δ_i . The following lemma will be vital for the necessary estimates.

Lemma 4.2. *Let $D \in \text{Div}(\Sigma)$ and let $X_0 \subset \Sigma$ be a zero-dimensional scheme such that there exist curves $\Delta_1, \dots, \Delta_m \subset \Sigma$ and zero-dimensional schemes $X_i^0 \subseteq X_{i-1}$ for $i = 1, \dots, m$, where $X_i = X_{i-1} : \Delta_i$ for $i = 1, \dots, m$, such that (a)–(f) in Lemma 4.1 are fulfilled.*

Then:

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)) &\leq \sum_{i=1}^m h^1\left(\Delta_i, \mathcal{J}_{X_{i-1}\cap\Delta_i/\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right)\right) \\ &\leq \sum_{i=1}^m \left(1 + \deg(X_{i-1} \cap \Delta_i) - \deg(X_i^0)\right) \\ &\leq \sum_{i=1}^m \left(\Delta_i \cdot (K_\Sigma + \sum_{k=1}^i \Delta_k) + 1\right). \end{aligned}$$

Proof. Throughout the proof we use the following notation:

$$\mathcal{G}_i = \mathcal{J}_{X_{i-1}\cap\Delta_i/\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right) \quad \text{and} \quad \mathcal{G}_i^0 = \mathcal{J}_{X_i^0/\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right)$$

for $i = 1, \dots, m$, and for $i = 0, \dots, m$,

$$\mathcal{F}_i = \mathcal{J}_{X_i/\Sigma}\left(D - \sum_{k=1}^i \Delta_k\right).$$

Since $X_{i+1} = X_i : \Delta_{i+1}$, we have the following short exact sequence:

$$(4.10) \quad 0 \longrightarrow \mathcal{F}_{i+1} \xrightarrow{\cdot\Delta_{i+1}} \mathcal{F}_i \longrightarrow \mathcal{G}_{i+1} \longrightarrow 0$$

for $i = 0, \dots, m-1$ and the corresponding long exact cohomology sequence

$$(4.11) \quad \begin{aligned} 0 \longrightarrow H^0(\Sigma, \mathcal{F}_{i+1}) \longrightarrow H^0(\Sigma, \mathcal{F}_i) \longrightarrow H^0(\Sigma, \mathcal{G}_{i+1}) \longrightarrow H^1(\Sigma, \mathcal{F}_{i+1}) \\ \downarrow \\ 0 = H^2(\Sigma, \mathcal{G}_{i+1}) \longleftarrow H^2(\Sigma, \mathcal{F}_i) \longleftarrow H^2(\Sigma, \mathcal{F}_{i+1}) \longleftarrow H^1(\Sigma, \mathcal{G}_{i+1}) \longleftarrow H^1(\Sigma, \mathcal{F}_i) \end{aligned}$$

Step 1: $h^1(\Sigma, \mathcal{F}_i) \leq \sum_{j=i+1}^m h^1(\Sigma, \mathcal{G}_j)$ for $i = 0, \dots, m-1$.

We prove the claim by descending induction on i . From (4.11) we deduce

$$0 = H^1(\Sigma, \mathcal{F}_m) \longrightarrow H^1(\Sigma, \mathcal{F}_{m-1}) \longrightarrow H^1(\Sigma, \mathcal{G}_m),$$

which implies $h^1(\Sigma, \mathcal{F}_{m-1}) \leq h^1(\Sigma, \mathcal{G}_m)$ and thus proves the case $i = m-1$.

We may therefore assume that $1 \leq i \leq m-2$. Once more from (4.11) we deduce

$$a = h^0(\Sigma, \mathcal{F}_{i+1}) - h^0(\Sigma, \mathcal{F}_i) + h^0(\Sigma, \mathcal{G}_{i+1}) \geq 0$$

and

$$b = h^2(\Sigma, \mathcal{F}_{i+1}) - h^2(\Sigma, \mathcal{F}_i) \geq 0,$$

and finally,

$$\begin{aligned} h^1(\Sigma, \mathcal{F}_i) &= h^1(\Sigma, \mathcal{G}_{i+1}) + h^1(\Sigma, \mathcal{F}_{i+1}) - a - b \leq h^1(\Sigma, \mathcal{G}_{i+1}) + h^1(\Sigma, \mathcal{F}_{i+1}) \\ &\leq_{\text{Ind.}} h^1(\Sigma, \mathcal{G}_{i+1}) + \sum_{j=i+2}^m h^1(\Sigma, \mathcal{G}_j) = \sum_{j=i+1}^m h^1(\Sigma, \mathcal{G}_j). \end{aligned}$$

Step 2: $h^1(\Delta_i, \mathcal{G}_i) = h^0(\Delta_i, \mathcal{G}_i) - \chi\left(\mathcal{O}_{\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right)\right) + \deg(X_{i-1} \cap \Delta_i)$.

We consider the exact sequence

$$0 \longrightarrow \mathcal{G}_i \longrightarrow \mathcal{O}_{\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right) \longrightarrow \mathcal{O}_{X_{i-1}\cap\Delta_i/\Delta_i}\left(D - \sum_{k=1}^{i-1} \Delta_k\right) \longrightarrow 0.$$

The result then follows from the long exact cohomology sequence.

Step 3: $h^0(\Delta_i, \mathcal{G}_i^0) - \chi\left(\mathcal{O}_{\Delta_i}(D - \sum_{k=1}^{i-1} \Delta_k)\right) = h^1(\Delta_i, \mathcal{G}_i^0) - \deg(X_i^0)$.

This follows analogously, replacing X_{i-1} by X_i^0 , since $X_i^0 = X_i^0 \cap \Delta_i$.

Step 4: $h^1(\Delta_i, \mathcal{G}_i^0) \leq h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) = 1$.

Note that $X_i^0 : \Delta_i = \emptyset$, and hence $\mathcal{J}_{X_i^0:\Delta_i/\Sigma} = \mathcal{O}_{\Sigma}$. We thus have the following short exact sequence:

$$(4.12) \quad 0 \longrightarrow \mathcal{O}_{\Sigma}\left(D - \sum_{k=1}^i \Delta_k\right) \xrightarrow{\cdot \Delta_i} \mathcal{J}_{X_i^0/\Sigma}\left(D - \sum_{k=1}^{i-1} \Delta_k\right) \longrightarrow \mathcal{G}_i^0 \longrightarrow 0.$$

By assumption (f) the divisor $D - K_{\Sigma} - \sum_{k=1}^i \Delta_k$ is big and nef and hence

$$0 = h^0\left(\Sigma, \mathcal{O}_{\Sigma}(-D + K_{\Sigma} + \sum_{k=1}^i \Delta_k)\right) = h^2\left(\Sigma, \mathcal{O}_{\Sigma}(D - \sum_{k=1}^i \Delta_k)\right).$$

Thus the long exact cohomology sequence of (4.12) gives

$$H^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right) \longrightarrow H^1(\Delta_i, \mathcal{G}_i^0) \longrightarrow 0$$

and

$$h^1(\Delta_i, \mathcal{G}_i^0) \leq h^1\left(\Sigma, \mathcal{J}_{X_i^0/\Sigma}(D - \sum_{k=1}^{i-1} \Delta_k)\right).$$

However, by assumption (b) the latter is just one.

Step 5: $h^1(\Delta_i, \mathcal{G}_i) \leq 1 + \deg(X_{i-1} \cap \Delta_i) - \deg(X_i^0)$.

We note that $\mathcal{G}_i \hookrightarrow \mathcal{G}_i^0$, and thus $h^0(\Delta_i, \mathcal{G}_i) \leq h^0(\Delta_i, \mathcal{G}_i^0)$. But then

$$\begin{aligned} h^1(\Delta_i, \mathcal{G}_i) &\leq_{\text{Step 2/3}} h^1(\Delta_i, \mathcal{G}_i^0) - \deg(X_i^0) + \deg(X_{i-1} \cap \Delta_i) \\ &\leq_{\text{Step 4}} 1 - \deg(X_i^0) + \deg(X_{i-1} \cap \Delta_i). \end{aligned}$$

Step 6: Finish the proof.

Taking into account that $h^1(\Sigma, \mathcal{G}_i) = h^1(\Delta_i, \mathcal{G}_i)$, since \mathcal{G}_i is concentrated on Δ_i , the first inequality follows from Step 1, while the second inequality is a consequence of Step 5, and the last inequality follows from assumption (c). \square

In Lemmata 4.3, 4.4 and 4.6 we consider special classes of surfaces which allow us to do the necessary estimates in order to finally derive

$$\sum_{i=1}^m (\#X_i^0 - \dim |\Delta_i|_l) > h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)).$$

We first consider surfaces with Picard number one.

Lemma 4.3. *Let Σ be a surface such that*

- (i) $\text{NS}(\Sigma) = L \cdot \mathbb{Z}$ and L is ample, and
- (ii) $h^1(\Sigma, C) = 0$, whenever C is effective.

Let $D \in \text{Div}(\Sigma)$ and let $X_0 \subset \Sigma$ be a zero-dimensional scheme satisfying (0)–(3) from Lemma 4.1 and

$$(4) \quad \sum_{z \in \Sigma} (\deg(X_{0,z}))^2 < \gamma \cdot (D - K_{\Sigma})^2, \text{ where } \gamma = \frac{(1 + \sqrt{1 - 4\beta})^2 \cdot L^2}{4 \cdot \chi(\mathcal{O}_{\Sigma}) + \max\{0, 2 \cdot K_{\Sigma} \cdot L\} + 6 \cdot L^2}.$$

Then, using the notation of Lemma 4.1 and setting $X_S = \bigcup_{i=1}^m X_i^0$,

$$h^1(\Sigma, \mathcal{J}_{X_0/\Sigma}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) < \#X_S.$$

Proof. We fix the following notation:

$$D \sim_a d \cdot L, \quad K_\Sigma \sim_a \kappa \cdot L, \quad \Delta_i \sim_a \delta_i \cdot L, \quad \text{and } l = \sqrt{L^2} > 0.$$

Furthermore, we have $\gamma = \frac{(1+\sqrt{1-4\beta})^2}{4\alpha}$, where

$$\alpha = \frac{4 \cdot \chi(\mathcal{O}_\Sigma) + \max\{0, 2 \cdot K_\Sigma \cdot L\} + 6 \cdot L^2}{4 \cdot L^2} = \begin{cases} \frac{\chi(\mathcal{O}_\Sigma)}{l^2} + \frac{\kappa+3}{2}, & \text{if } \kappa \geq 0, \\ \frac{\chi(\mathcal{O}_\Sigma)}{l^2} + \frac{3}{2}, & \text{if } \kappa < 0. \end{cases}$$

Step 1: By (i) Σ satisfies assumption (*) of Lemma 4.1.

Step 2: $\sum_{i=1}^m \delta_i \cdot l \leq \frac{(d-\kappa) \cdot l}{2} - \sqrt{\frac{(d-\kappa)^2 \cdot l^2}{4} - \deg(X_S)}$, by (4.1).

Step 3: $h^1(\Sigma, \mathcal{J}_{X_0}(D)) \leq (\kappa \cdot \sum_{i=1}^m \delta_i) \cdot l^2 + \frac{1}{2} \left(\left(\sum_{i=1}^m \delta_i \right)^2 + \sum_{i=1}^m \delta_i^2 \right) \cdot l^2 + m$.
By Lemma 4.2 we know that

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X_0}(D)) &\leq \sum_{i=1}^m \left(\Delta_i \cdot (K_\Sigma + \sum_{k=1}^i \Delta_k) + 1 \right) \\ &= \left(\kappa \cdot \sum_{i=1}^m \delta_i \right) \cdot l^2 + \frac{1}{2} \left(\left(\sum_{i=1}^m \delta_i \right)^2 + \sum_{i=1}^m \delta_i^2 \right) \cdot l^2 + m. \end{aligned}$$

Step 4: $\sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) \leq m \cdot (\chi(\mathcal{O}_\Sigma) - 1) + \frac{l^2}{2} \cdot \sum_{i=1}^m \delta_i^2 - \frac{\kappa \cdot l^2}{2} \cdot \sum_{i=1}^m \delta_i$.
Since Δ_i is effective by (ii), $h^1(\Sigma, \Delta_i) = 0$. Hence by Riemann–Roch

$$\begin{aligned} \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) &\leq -m + m \cdot \chi(\mathcal{O}_\Sigma) + \frac{1}{2} \sum_{i=1}^m (\Delta_i^2 - K_\Sigma \cdot \Delta_i) \\ &= m \cdot (\chi(\mathcal{O}_\Sigma) - 1) + \frac{l^2}{2} \cdot \sum_{i=1}^m \delta_i^2 - \frac{\kappa \cdot l^2}{2} \cdot \sum_{i=1}^m \delta_i. \end{aligned}$$

Step 5: Finish the proof.

In the following consideration we use that $\deg(X_S) \leq \deg(X_0) \leq \beta \cdot (d - \kappa)^2 \cdot l^2$:

$$\begin{aligned} & h^1(\Sigma, \mathcal{I}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) \\ & \leq_{\text{Step 3 / 4}} m \cdot \chi(\mathcal{O}_\Sigma) + l^2 \cdot \sum_{i=1}^m \delta_i^2 + \frac{\kappa \cdot l^2}{2} \cdot \sum_{i=1}^m \delta_i + \frac{l^2}{2} \cdot \left(\sum_{i=1}^m \delta_i \right)^2 \\ & \leq \alpha \cdot \left(l \cdot \sum_{i=1}^m \delta_i \right)^2 \leq_{\text{Step 2}} \alpha \cdot \left(\frac{(d - \kappa) \cdot l}{2} - \sqrt{\frac{(d - \kappa)^2 \cdot l^2}{4} - \deg(X_S)} \right)^2 \\ & \leq \alpha \cdot \left(\frac{\frac{(d - \kappa)^2 \cdot l^2}{4} - \left(\frac{(d - \kappa)^2 \cdot l^2}{4} - \deg(X_S) \right)}{\frac{(d - \kappa) \cdot l}{2} + \sqrt{\frac{(d - \kappa)^2 \cdot l^2}{4} - \deg(X_S)}} \right)^2 \\ & = \alpha \cdot \left(\frac{2 \cdot \deg(X_S)}{(d - \kappa) \cdot l + \sqrt{(d - \kappa)^2 \cdot l^2 - 4 \cdot \deg(X_S)}} \right)^2 \\ & \leq \frac{4\alpha}{(1 + \sqrt{1 - 4\beta})^2 \cdot (d - \kappa)^2 \cdot l^2} \cdot (\deg(X_S))^2 \\ & = \frac{1}{\gamma \cdot (D - K_\Sigma)^2} \cdot \left(\sum_{z \in \Sigma} \deg(X_{S,z}) \right)^2 \\ & \leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{S,z})^2 \\ & \leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{0,z})^2 \stackrel{(4)}{<} \#X_S. \end{aligned}$$

□

The second class of surfaces that we consider are products of curves.

Lemma 4.4. *Let C_1 and C_2 be two smooth projective curves of genera g_1 and g_2 , respectively, with $g_1 \geq g_2 \geq 0$, such that for $\Sigma = C_1 \times C_2$, the Néron–Severi group is $\text{NS}(\Sigma) = C_1\mathbb{Z} \oplus C_2\mathbb{Z}$, and let $D \in \text{Div}(\Sigma)$ be such that $D \sim_a aC_1 + bC_2$ with $a > \max\{2g_2 - 2, 2 - 2g_2\}$ and $b > \max\{2g_1 - 2, 2 - 2g_1\}$. Suppose moreover that $X_0 \subset \Sigma$ is a zero-dimensional scheme satisfying (1)–(3) from Lemma 4.1 and*

$$(4) \sum_{z \in \Sigma} (\deg(X_{0,z}))^2 < \gamma \cdot (D - K_\Sigma)^2,$$

where γ may be taken from the table in Theorem 2.7.

Then, using the notation of Lemma 4.1 and setting $X_S = \bigcup_{i=1}^m X_i^0$,

$$h^1(\Sigma, \mathcal{I}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) < \#X_S.$$

Proof. Then $K_\Sigma \sim_a (2g_2 - 2) \cdot C_1 + (2g_1 - 2) \cdot C_2$ and we fix the notation:

$$\Delta_i \sim_a a_i C_1 + b_i C_2, \quad \kappa_1 = a - 2g_2 + 2 \quad \text{and} \quad \kappa_2 = b - 2g_1 + 2.$$

Step 1: Σ satisfies assumption (*) of Lemma 4.1. Moreover, due to the assumptions on a and b we know that $D - K_\Sigma$ is ample and $D + K_\Sigma$ is nef, i.e., (0) in Lemma 4.1 is fulfilled as well.

Step 2a: $\left(\frac{\kappa_1}{4}\right) \cdot \sum_{i=1}^m b_i + \left(\frac{\kappa_2}{4}\right) \cdot \sum_{i=1}^m a_i \leq \deg(X_S).$

Let us first notice that the strict inequality “ $<$ ” in Lemma 4.1 (e) for ample divisors H comes down to “ \leq ” for nef divisors H . We may apply this for $H = C_1$ and $H = C_2$ and deduce the following inequalities:

$$(4.13) \quad 0 \leq \left(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i \right) \cdot C_1 = \kappa_2 - \sum_{k=1}^i b_k - b_i$$

and

$$(4.14) \quad 0 \leq \left(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i \right) \cdot C_2 = \kappa_1 - \sum_{k=1}^i a_k - a_i.$$

For the following consideration we choose $i_0, j_0 \in \{1, \dots, m\}$ such that $a_{i_0} \geq a_i$ for all $i = 1, \dots, m$ and $b_{j_0} \geq b_j$ for all $j = 1, \dots, m$. Then

$$(4.15) \quad \kappa_1 \geq 2a_{i_0} \quad \text{and} \quad \kappa_2 \geq 2b_{j_0}$$

for all $i, j = 1, \dots, m$; finally (4.13)–(4.15) lead to

$$\begin{aligned} \deg(X_S) &= \sum_{i=1}^m \deg(X_i^0) \geq_{\text{Lemma 4.1 (c)}} \sum_{i=1}^m \left(D - K_\Sigma - \sum_{k=1}^i \Delta_k \right) \cdot \Delta_i \\ &= \kappa_1 \sum_{i=1}^m b_i + \kappa_2 \sum_{i=1}^m a_i - \sum_{i=1}^m a_i b_i - \sum_{i=1}^m a_i \sum_{i=1}^m b_i \\ &\geq \frac{\kappa_1}{2} \sum_{i=1}^m b_i + \frac{\kappa_2}{2} \sum_{i=1}^m a_i + \frac{a_m}{2} \sum_{i=1}^m b_i + \frac{b_m}{2} \sum_{i=1}^m a_i - \sum_{i=1}^m a_i b_i \\ &\geq \frac{\kappa_1}{4} \sum_{i=1}^m b_i + \frac{\kappa_2}{4} \sum_{i=1}^m a_i. \end{aligned}$$

Step 2b: $\sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i \leq \frac{8}{(D-K_\Sigma)^2} \cdot (\deg(X_S))^2.$

Using Step 2a we deduce

$$\begin{aligned} (\deg(X_S))^2 &> \left(\frac{\kappa_2}{4} \cdot \sum_{i=1}^m a_i + \frac{\kappa_1}{4} \cdot \sum_{i=1}^m b_i \right)^2 \\ &\geq \frac{4 \cdot \kappa_1 \cdot \kappa_2}{16} \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i \\ &= \frac{(D-K_\Sigma)^2}{8} \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i. \end{aligned}$$

Step 2c: $\sum_{i=1}^m a_i \leq \begin{cases} \frac{2\alpha}{(D-K_\Sigma)^2} \cdot (\deg(X_S))^2, & \text{if } \sum_{i=1}^m b_i = 0, \\ \frac{8}{(D-K_\Sigma)^2} \cdot (\deg(X_S))^2, & \text{otherwise.} \end{cases}$

If $\sum_{i=1}^m b_i = 0$, then the same consideration as in Step 2a shows that

$$\deg(X_S) \geq \kappa_2 \cdot \sum_{i=1}^m a_i > 0,$$

and thus

$$\frac{(D-K_\Sigma)^2}{2\alpha} \cdot \sum_{i=1}^m a_i \leq \kappa_2^2 \cdot \left(\sum_{i=1}^m a_i \right)^2 \leq (\deg(X_S))^2.$$

If $\sum_{i=1}^m b_i \neq 0$, then we are done by Step 2b.

$$\text{Step 2d: } \sum_{i=1}^m b_i \leq \begin{cases} \frac{2}{\alpha \cdot (D-K_\Sigma)^2} \cdot (\deg(X_S))^2, & \text{if } \sum_{i=1}^m a_i = 0, \\ \frac{8}{(D-K_\Sigma)^2} \cdot (\deg(X_S))^2, & \text{otherwise.} \end{cases}$$

This is proved in the same way as Step 2c.

$$\text{Step 3: } h^1(\Sigma, \mathcal{J}_{X_0}(D)) \leq 2 \sum_{i=1}^m a_i \sum_{i=1}^m b_i + (2g_1 - 2) \sum_{i=1}^m a_i + (2g_2 - 2) \sum_{i=1}^m b_i + m.$$

The following sequence of inequalities is due to Lemma 4.2 and the fact that $\Delta_i \cdot \Delta_j \geq 0$ for any $i, j \in \{1, \dots, m\}$:

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X_0}(D)) &\leq \sum_{i=1}^m \left(\Delta_i \cdot (K_\Sigma + \sum_{k=1}^i \Delta_k) + 1 \right) \\ &\leq K_\Sigma \cdot \sum_{i=1}^m \Delta_i + \left(\sum_{i=1}^m \Delta_i \right)^2 + m \\ &= (2g_1 - 2) \cdot \sum_{i=1}^m a_i + (2g_2 - 2) \cdot \sum_{i=1}^m b_i + 2 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + m. \end{aligned}$$

Step 4: We find the estimate $\sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) \leq \beta$, where

$$\beta = \begin{cases} \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m b_i, & \text{if } g_1 = 1, g_2 = 0, \\ \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i - m, & \text{if } g_1 = 1, g_2 = 1, \exists i_0 : a_{i_0} b_{i_0} > 0, \\ \sum_{i=1}^m a_i + \sum_{i=1}^m b_i - m, & \text{if } g_1 = 1, g_2 = 1, \forall i : a_i b_i = 0, \\ \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m a_i + \sum_{i=1}^m b_i, & \text{otherwise.} \end{cases}$$

In general $h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) \leq a_i b_i + a_i + b_i + 1$, whereas if $g_1 = 1, g_2 = 0$, we have $h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) = a_i b_i + b_i + 1$. It thus only remains to consider the case $g_1 = g_2 = 1$, where we get

$$\sum_{i=1}^m h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) = \sum_{a_i, b_i > 0} a_i b_i + \sum_{a_i = 0} b_i + \sum_{b_i = 0} a_i.$$

If always either a_i or b_i is zero, we are done. Otherwise there exists some $i_0 \in \{1, \dots, m\}$ such that $a_{i_0} \neq 0 \neq b_{i_0}$. Then looking at the right-hand side we see

$$\sum_{i=1}^m h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) \leq \sum_{a_i, b_i > 0} a_i b_i + a_{i_0} \cdot \sum_{a_i = 0} b_i + b_{i_0} \cdot \sum_{b_i = 0} a_i \leq \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i.$$

Step 5: Finish the proof.

Using Step 3 and Step 4, and taking $m \leq \sum_{i=1}^m a_i + b_i$ into account, we get $h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) \leq \beta'$, where β' may be chosen as

$$\beta' = \begin{cases} 3 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i, & \text{if } g_1 = 0, g_2 = 0, \\ 3 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m a_i, & \text{if } g_1 = 1, g_2 = 0, \\ 3 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + 2g_1 \cdot \sum_{i=1}^m a_i + 2g_2 \cdot \sum_{i=1}^m b_i, & \text{if } g_1 \geq 2, g_2 \geq 0. \end{cases}$$

For the case $g_1 = g_2 = 1$ we take a closer look. We find at once the following upper bounds β'' for $h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right)$:

$$\beta'' = \begin{cases} 3 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i, & \text{if } \exists i_0 : a_{i_0} b_{i_0} \neq 0, \\ 2 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m a_i + \sum_{i=1}^m b_i, & \text{if } \forall i : a_i b_i = 0. \end{cases}$$

Considering now the cases $\sum_{i=1}^m a_i \neq 0 \neq \sum_{i=1}^m b_i$, $\sum_{i=1}^m a_i = 0$ and $\sum_{i=1}^m b_i = 0$, we can replace these by

$$\beta'' \leq \beta' = \begin{cases} 4 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i, & \text{if } \sum_{i=1}^m a_i \neq 0 \neq \sum_{i=1}^m b_i, \\ \sum_{i=1}^m a_i, & \text{if } \sum_{i=1}^m b_i = 0, \\ \sum_{i=1}^m b_i, & \text{if } \sum_{i=1}^m a_i = 0. \end{cases}$$

Applying now the results of Step 2 in all cases we get

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) &\leq \beta' \leq \frac{1}{\gamma \cdot (D - K_\Sigma)^2} \cdot (\deg(X_S))^2 \\ &= \frac{1}{\gamma \cdot (D - K_\Sigma)^2} \cdot \left(\sum_{z \in \Sigma} \deg(X_{S,z}) \right)^2 \leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{S,z})^2 \\ &\leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{0,z})^2 <_{(4)} \#X_S. \end{aligned}$$

□

Remark 4.5. Lemma 4.4, and hence Theorem 2.7 could easily be generalised to other surfaces Σ with irreducible curves $C_1, C_2 \subset \Sigma$ such that $\text{NS}(\Sigma) = C_1\mathbb{Z} \oplus C_2\mathbb{Z}$ with intersection matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ once we have an estimate similar to

$$h^0(\Sigma, aC_1 + bC_2) \leq ab + a + b + 1$$

for an effective divisor $aC_1 + bC_2$.

With a number of small modifications, we are even able to adapt it in the following lemma in the case of geometrically ruled surfaces with non-positive invariant e although the intersection pairing looks more complicated.

The problem with arbitrary geometrically ruled surfaces is the existence of the section with negative self-intersection, once the invariant $e > 0$, since then the proof of Lemma 4.1 no longer works.

In the following lemma we use the notation of Subsection 2.3.

Lemma 4.6. *Let $\pi : \Sigma \rightarrow C$ be a geometrically ruled surface with invariant $e \leq 0$ and $g = g(C)$, and let $D \in \text{Div}(\Sigma)$ be such that $D \sim_a aC_0 + bF$ with $a \geq 2$, $b > 2g - 2 + \frac{ae}{2}$, and if $g = 0$, then $b \geq 2$. Suppose moreover that $X_0 \subset \Sigma$ is a zero-dimensional scheme satisfying (1)–(3) from Lemma 4.1 and*

$$(4) \sum_{z \in \Sigma} (\deg(X_{0,z}))^2 < \gamma \cdot (D - K_\Sigma)^2,$$

where γ may be taken from the table in Theorem 2.8.

Then, using the notation of Lemma 4.1 and setting $X_S = \bigcup_{i=1}^m X_i^0$,

$$h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) < \#X_S.$$

Proof. Remember that the Néron–Severi group of Σ is generated by a section C_0 of π and a fibre F with intersection pairing given by $\begin{pmatrix} -e & 1 \\ 1 & 0 \end{pmatrix}$. Then $K_\Sigma \sim_a -2C_0 + (2g - 2 - e) \cdot F$ and we fix the notation:

$$\Delta_i \sim_a a_i C_0 + b'_i F.$$

Note that then

$$a_i \geq 0 \quad \text{and} \quad b_i := b'_i - \frac{e}{2} a_i \geq 0.$$

Finally, we set $\kappa_1 = a + 2$ and $\kappa_2 = b + 2 - 2g - \frac{ae}{2}$ and get

$$(4.16) \quad (D - K_\Sigma)^2 = -e \cdot (a + 2)^2 + 2 \cdot (a + 2) \cdot (b + 2 + e - 2g) = 2 \cdot \kappa_1 \cdot \kappa_2.$$

Replacing the equations (4.13) and (4.14) by

$$(4.17) \quad 0 \leq \left(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i \right) \cdot (C_0 + \frac{e}{2} F) = \kappa_2 - \sum_{k=1}^i b_k - b_i$$

and

$$(4.18) \quad 0 \leq \left(D - K_\Sigma - \sum_{k=1}^i \Delta_k - \Delta_i \right) \cdot F = \kappa_1 - \sum_{k=1}^i a_k - a_i,$$

the assertions of Step 1 to Step 2c in the proof of Lemma 4.4 remain literally true.

Step 2d: $\left(\sum_{i=1}^m a_i \right)^2 \leq \frac{32\alpha}{(D-K_\Sigma)^2} (\deg(X_S))^2$ and $\left(\sum_{i=1}^m b_i \right)^2 \leq \frac{32}{\alpha \cdot (D-K_\Sigma)^2} (\deg(X_S))^2$.

This follows from the following inequality with the aid of Step 2a and (4.16):

$$\begin{aligned} (\deg(X_S))^2 &\geq \left(\frac{\kappa_2}{4} \cdot \sum_{i=1}^m a_i \right)^2 + \left(\frac{\kappa_1}{4} \cdot \sum_{i=1}^m b_i \right)^2 \\ &\geq \frac{2 \cdot \kappa_1 \cdot \kappa_2}{32\alpha} \cdot \left(\sum_{i=1}^m a_i \right)^2 + \frac{2 \cdot \kappa_1 \cdot \kappa_2 \cdot \alpha}{32} \cdot \left(\sum_{i=1}^m b_i \right)^2. \end{aligned}$$

Step 3: $h^1(\Sigma, \mathcal{J}_{X_0}(D)) \leq 2 \cdot \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + (2g - 2) \cdot \sum_{i=1}^m a_i - 2 \cdot \sum_{i=1}^m b_i + m$ is proved as Step 3 in Lemma 4.4.

Step 4a: If $e = 0$, we find the estimate

$$\sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) \leq \begin{cases} \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m b_i - m, & \text{if } g = 1, \sum_{i=1}^m b_i \neq 0, \\ \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m b_i = 0, & \text{if } g = 1, \sum_{i=1}^m b_i = 0, \\ \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m a_i + \sum_{i=1}^m b_i, & \text{for } g \text{ arbitrary.} \end{cases}$$

We note that in this case $b'_i = b_i$ and that $b_i = 0$ thus implies $a_i > 0$. But then

$$h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) \leq \begin{cases} a_i b_i + b_i, & \text{if } g = 1, b_i > 0, \\ a_i b_i + b_i + 1 = 1, & \text{if } g = 1, b_i = 0, \\ a_i b_i + a_i + b_i + 1, & \text{otherwise.} \end{cases}$$

The results for g arbitrary, respectively, $g = 1$ and $\sum_{i=1}^m b_i = 0$ thus follow right away. If, however, some $b_{i_0} > 0$, then $\sum_{i \neq j} a_i b_j \geq b_{i_0} \sum_{i \neq i_0} a_i \geq \#\{b_i \mid b_i = 0\}$

and hence

$$\begin{aligned} h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) &\leq \sum_{i=1}^m a_i b_i + \sum_{i=1}^m b_i + \#\{b_i \mid b_i = 0\} \\ &= \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m b_i + \#\{b_i \mid b_i = 0\} - \sum_{i \neq j} a_i b_j \leq \sum_{i=1}^m a_i \cdot \sum_{i=1}^m b_i + \sum_{i=1}^m b_i. \end{aligned}$$

Step 4b: If $e < 0$, we give several upper bounds for $\beta = \sum_{i=1}^m (h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1)$:

$$\beta \leq \begin{cases} \frac{1}{2} \sum_{i=1}^m a_i \sum_{i=1}^m b_i + \frac{1}{2} \left(\sum_{i=1}^m b_i \right)^2 + \frac{1}{8} \left(\sum_{i=1}^m a_i \right)^2 + \frac{1}{4} \sum_{i=1}^m a_i + \frac{1}{2} \sum_{i=1}^m b_i, & \text{if } g = 1, \\ \sum_{i=1}^m a_i \sum_{i=1}^m b_i + \sum_{i=1}^m a_i + \sum_{i=1}^m b_i - \frac{9e}{32} \left(\sum_{i=1}^m a_i \right)^2, & \text{for } g \text{ arbitrary,} \\ \frac{1}{4} \sum_{i=1}^m a_i \sum_{i=1}^m b_i + \sum_{i=1}^m a_i + \sum_{i=1}^m b_i - \frac{9e}{32} \left(\sum_{i=1}^m a_i \right)^2 - \frac{1}{2e} \left(\sum_{i=1}^m b_i \right)^2, & g \text{ arbitrary.} \end{cases}$$

If g is arbitrary, the claim follows since a thorough investigation leads to

$$h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) \leq a_i b_i + a_i + b_i + 1 - \frac{9e}{32} \cdot a_i^2$$

and

$$h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) \leq \frac{1}{4} \cdot a_i b_i + a_i + b_i + 1 - \frac{9e}{32} \cdot a_i^2 - \frac{1}{2e} \cdot b_i^2.$$

If $g = 1$, then $e = -1$ and $b = b' + \frac{a}{2}$, and we are done since

$$\begin{aligned} h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) &\leq a_i b'_i + b'_i + 1 + \frac{a_i(a_i+1)}{2} + \frac{b'_i(b'_i-1)}{2} \\ &= \frac{1}{2} \cdot a_i b_i + \frac{1}{2} \cdot b_i^2 + \frac{1}{8} \cdot a_i^2 + \frac{1}{4} \cdot a_i + \frac{1}{2} \cdot b_i + 1. \end{aligned}$$

Step 5: In this last step we gather the information from the previous investigations and finish the proof considering a bunch of different cases.

Using Step 3 and Step 4 and taking $\sum_{i=1}^m a_i + b_i \leq m$ into account, we get the following upper bounds for $\beta' = h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m (h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1)$:

$$\beta' \leq \begin{cases} 3 \sum_{i=1}^m a_i \sum_{i=1}^m b_i + 2g \sum_{i=1}^m a_i, & \text{if } e = 0, \\ 3 \sum_{i=1}^m a_i \sum_{i=1}^m b_i + 2g \sum_{i=1}^m a_i - \frac{9e}{32} \left(\sum_{i=1}^m a_i \right)^2, & \text{if } e < 0, \\ \frac{9}{4} \sum_{i=1}^m a_i \sum_{i=1}^m b_i + 2g \sum_{i=1}^m a_i - \frac{9e}{32} \left(\sum_{i=1}^m a_i \right)^2 - \frac{1}{2e} \left(\sum_{i=1}^m b_i \right)^2, & \text{if } e < 0, \\ 3 \sum_{i=1}^m a_i \sum_{i=1}^m b_i, & \text{if } e = 0, g = 1, \sum_{i=1}^m b_i \neq 0, \\ m \leq \sum_{i=1}^m a_i, & \text{if } e = 0, g = 1, \sum_{i=1}^m b_i = 0, \\ \frac{5}{2} \sum_{i=1}^m a_i \sum_{i=1}^m b_i + \frac{1}{2} \left(\sum_{i=1}^m b_i \right)^2 + \frac{1}{8} \left(\sum_{i=1}^m a_i \right)^2 + \frac{5}{4} \sum_{i=1}^m a_i, & \text{if } e < 0, g = 1. \end{cases}$$

Applying now Steps 2b–2d we end up with $\frac{\beta' \cdot (D - K_\Sigma)^2}{(\deg(X_S))^2} \leq \gamma$. We thus finally get

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) &= \beta' \leq \frac{1}{\gamma \cdot (D - K_\Sigma)^2} \cdot (\deg(X_S))^2 \\ &= \frac{1}{\gamma \cdot (D - K_\Sigma)^2} \cdot \left(\sum_{z \in \Sigma} \deg(X_{S,z}) \right)^2 \leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{S,z})^2 \\ &\leq \frac{\#X_S}{\gamma \cdot (D - K_\Sigma)^2} \cdot \sum_{z \in \Sigma} \deg(X_{0,z})^2 <_{(4)} \#X_S. \end{aligned}$$

□

It remains to show that the inequality that we derived cannot hold.

Lemma 4.7. *Let $D \in \text{Div}(\Sigma)$, let S_1, \dots, S_r be pairwise distinct topological or analytical singularity types and let $k_1, \dots, k_r \in \mathbb{N} \setminus \{0\}$. Suppose that $V_{|D|}^{irr,reg}(k_1 S_1, \dots, k_r S_r)$ is nonempty.*

Then there exists no curve $C \in V_{|D|}^{irr}(k_1 S_1, \dots, k_r S_r) \setminus \overline{V_{|D|}^{irr,reg}(k_1 S_1, \dots, k_r S_r)}$ such that for the zero-dimensional scheme $X_0 = X(C)$ there exist curves $\Delta_1, \dots, \Delta_m \subset \Sigma$ and zero-dimensional locally complete intersections $X_i^0 \subseteq X_{i-1}$ for $i = 1, \dots, m$, where $X_i = X_{i-1} : \Delta_i$ for $i = 1, \dots, m$, such that $X_S = \bigcup_{i=1}^m X_i^0$ satisfies

$$(4.19) \quad h^1(\Sigma, \mathcal{J}_{X_0}(D)) + \sum_{i=1}^m \left(h^0(\Sigma, \mathcal{O}_\Sigma(\Delta_i)) - 1 \right) < \#X_S.$$

Proof. Throughout the proof we use the notation $V^{irr} = V_{|D|}^{irr}(k_1 S_1, \dots, k_r S_r)$ and $V^{irr,reg} = V_{|D|}^{irr,reg}(k_1 S_1, \dots, k_r S_r)$.

Suppose there exists a curve $C \in V^{irr} \setminus \overline{V^{irr,reg}}$ satisfying the assumption of the lemma, and let V^* be the irreducible component of V^{irr} containing C . Moreover, let $C_0 \in V^{irr,reg}$.

In the following we consider the morphism from Subsection 1.5,

$$\Psi = \Psi_{|D|}(k_1 S_1, \dots, k_r S_r) : V_{|D|}(k_1 S_1, \dots, k_r S_r) \rightarrow B(k_1 S_1, \dots, k_r S_r) = B.$$

Step 1: $h^0(\Sigma, \mathcal{J}_{X(C_0)/\Sigma}(D)) = h^0(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) - h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D))$.

By the choice of C_0 we have

$$0 = H^1(\Sigma, \mathcal{J}_{X^*(C_0)/\Sigma}(D)) \rightarrow H^1(\Sigma, \mathcal{O}_\Sigma(D)) \rightarrow H^1(\Sigma, \mathcal{O}_{X^*(C_0)}(D)) = 0,$$

and thus D is non-special, i.e., $h^1(\Sigma, \mathcal{O}_\Sigma(D)) = 0$. But then

$$h^0(\Sigma, \mathcal{J}_{X(C_0)/\Sigma}(D)) = h^0(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) - h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)).$$

Step 2: $h^1(\Sigma, \mathcal{J}_{X(C)}(D)) \geq \text{codim}_B(\Psi(V^*))$.

Suppose the contrary, that is, $\dim(\Psi(V^*)) < \dim(B) - h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D))$. Then by Step 1 and Theorem 3.1,

$$\begin{aligned} \dim(V^*) &\leq \dim(\Psi(V^*)) + \dim(\Psi^{-1}(\Psi(C))) \\ &< \dim(B) - h^1(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) + h^0(\Sigma, \mathcal{J}_{X(C)/\Sigma}(D)) - 1 \\ &= \dim(B) + h^0(\Sigma, \mathcal{J}_{X(C_0)/\Sigma}(D)) - 1 = \dim(V^{irr,reg}). \end{aligned}$$

However, any irreducible component of V^{irr} has at least the expected dimension $\dim(V^{irr,reg})$, which gives a contradiction.

Step 3: $\text{codim}_B(\Psi(V^*)) \geq \#X_S - \sum_{i=1}^m \dim |\Delta_i|_l$. The existence of the subschemes $X_i^0 \subseteq X(C) \cap \Delta_i$ imposes at least $\#X_i^0 - \dim |\Delta_i|_l$ conditions on $X(C)$ and increases thus the codimension of $\Psi(V^*)$ by this number.

Step 4: Collecting the results we derive the following contradiction:

$$\begin{aligned} h^1(\Sigma, \mathcal{J}_{X(C)}(D)) &\geq_{\text{Step 2}} \text{codim}_B(\Psi(V^*)) \\ &\geq_{\text{Step 3}} \#X_S - \sum_{i=1}^m \dim |\Delta_i|_l >_{(4.19)} h^1(\Sigma, \mathcal{J}_{X(C)}(D)). \end{aligned}$$

□

REFERENCES

- [AC83] Enrico Arbarello and Maurizio Cornalba, *A few remarks about the variety of irreducible plane curves of given degree and genus*, Ann. Sci. École Norm. Sup. **16** (1983), no. 3, 467–488. MR **86a**:14020
- [Bar93] Didier Barkats, *Irréductibilité des variétés des courbes planes à noeuds et à cusps*, Preprint no. 364, Univ. de Nice-Sophia-Antipolis, 1993.
- [Bog79] Fedor A. Bogomolov, *Holomorphic tensors and vector bundles on projective manifolds*, Math. USSR Izvestija **13** (1979), no. 3, 499–555. MR **80j**:14014
- [Bru99] Andrea Bruno, *Limit linear series and families of equisingular plane curves*, Preprint, 1999.
- [CC99] Luca Chiantini and Ciro Ciliberto, *On the Severi variety of surfaces in $\mathbb{P}_{\mathbb{C}}^3$* , J. Algebraic Geom. **8** (1999), 67–83. MR **2000f**:14082
- [Che01] Xi Chen, *Some remarks on the Severi varieties of surfaces in \mathbb{P}^3* , Int. J. Math. Math. Sci. **26** (2001), no. 1, 1–5. MR **2002f**:14011
- [DH88] Steven Diaz and Joe Harris, *Ideals associated to deformations of singular plane curves*, Trans. Amer. Math. Soc. **309** (1988), 433–468. MR **89m**:14003
- [Fla01] Flaminio Flamini, *Some results of regularity for Severi varieties of projective surfaces*, Comm. Algebra **29** (2001), no. 6, 2297–2311. MR **2002c**:14043
- [GK89] Gert-Martin Greuel and Ulrich Karras, *Families of varieties with prescribed singularities*, Compositio Math. **69** (1989), 83–110. MR **90d**:32037
- [GL96] Gert-Martin Greuel and Christoph Lossen, *Equianalytic and equisingular families of curves on surfaces*, Manuscripta Math. **91** (1996), 323–342. MR **98g**:14023
- [GLS97] Gert-Martin Greuel, Christoph Lossen, and Eugenii Shustin, *New asymptotics in the geometry of equisingular families of curves*, Internat. Math. Res. Notices **13** (1997), 595–611. MR **98g**:14039
- [GLS98a] Gert-Martin Greuel, Christoph Lossen, and Eugenii Shustin, *Geometry of families of nodal curves on the blown-up projective plane*, Trans. Amer. Math. Soc. **350** (1998), 251–274. MR **98j**:14034
- [GLS98b] Gert-Martin Greuel, Christoph Lossen, and Eugenii Shustin, *On the irreducibility of families of curves*, Unpublished Manuscript, 1998.
- [GLS98c] Gert-Martin Greuel, Christoph Lossen, and Eugenii Shustin, *Plane curves of minimal degree with prescribed singularities*, Inv. math. **133** (1998), 539–580. MR **99g**:14035
- [GLS00] Gert-Martin Greuel, Christoph Lossen, and Eugenii Shustin, *Castelnuovo function, zero-dimensional schemes, and singular plane curves*, J. Algebraic Geom. **9** (2000), no. 4, 663–710. MR **2001g**:14045
- [Har77] Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977. MR **57**:3116
- [Har85] Joe Harris, *On the Severi problem*, Inventiones Math. **84** (1985), 445–461. MR **87f**:14012
- [Kan89a] Pyung-Lyun Kang, *A note on the variety of plane curves with nodes and cusps*, Proc. Amer. Math. Soc. **106** (1989), no. 2, 309–312. MR **89k**:14046
- [Kan89b] Pyung-Lyun Kang, *On the variety of plane curves of degree d with δ nodes and k cusps*, Trans. Amer. Math. Soc. **316** (1989), no. 1, 165–192. MR **90g**:14014

- [KT02] Thomas Keilen and Ilya Tyomkin, *Existence of curves with prescribed topological singularities*, Trans. Amer. Math. Soc. **354** (2002), no. 5, 1837–1860, <http://www.mathematik.uni-kl.de/~keilen/download/KeilenTyomkin001/KeilenTyomkin001.ps.gz>. MR **2003a**:14041
- [Laz97] Robert Lazarsfeld, *Lectures on linear series*, Complex Algebraic Geometry (János Kollár, ed.), IAS/Park City Mathematics Series, no. 3, Amer. Math. Soc., Providence, RI, 1997, pp. 161–219. MR **98h**:14008
- [Los98] Christoph Lossen, *The geometry of equisingular and equianalytic families of curves on a surface*, Ph.D. Thesis, FB Mathematik, Universität Kaiserslautern, Aug. 1998, <http://www.mathematik.uni-kl.de/~lossen/download/Lossen002/Lossen002.ps.gz>.
- [Mil68] John Milnor, *Singular points of complex hypersurfaces*, Princeton University Press, Princeton, NJ, 1968. MR **39**:969
- [Ran89] Ziv Ran, *Families of plane curves and their limits: Enriques' conjecture and beyond*, Annals of Math. **130** (1989), no. 1, 121–157. MR **90e**:14024
- [Sev21] Francesco Severi, *Vorlesungen über Algebraische Geometrie*, Bibliotheca Mathematica Teubneriana, no. 32, Teubner, 1921.
- [Shu91a] Eugenii Shustin, *Geometry of discriminant and topology of algebraic curves*, Proc. Internat. Congress Math., Kyoto 1990 (Tokyo, Berlin, New York), vol. 1, Springer-Verlag, 1991. MR **93h**:14018
- [Shu91b] Eugenii Shustin, *On manifolds of singular algebraic curves*, Selecta Math. Sov. **10** (1991), 27–37. MR **91k**:00045
- [Shu94] Eugenii Shustin, *Smoothness and irreducibility of varieties of plane curves with nodes and cusps*, Bull. Soc. Math. France **122** (1994), 235–253. MR **95e**:14020
- [Shu96a] Eugenii Shustin, *Geometry of equisingular families of plane algebraic curves*, J. Alg. Geom. **5** (1996), 209–234. MR **97g**:14025
- [Shu96b] Eugenii Shustin, *Smoothness and irreducibility of varieties of algebraic curves with ordinary singularities*, Israel Math. Conf. Proc., no. 9, Amer. Math. Soc., 1996, pp. 393–416. MR **97c**:14028
- [Wal96] Charles T.Č. Wall, *Highly singular quintic curves*, Math. Proc. Cambridge Philos. Soc. **119** (1996), 257–277. MR **97b**:14058
- [Zar35] Oscar Zariski, *Algebraic surfaces*, vol. III, Ergebnisse der Mathematik und ihrer Grenzgebiete, no. 5, Springer, 1935. MR **57**:9695

UNIVERSITÄT KAISERSLAUTERN, FACHBEREICH MATHEMATIK, ERWIN-SCHRÖDINGER-STRASSE,
D-67663 KAISERSLAUTERN, GERMANY

E-mail address: keilen@mathematik.uni-kl.de

URL: <http://www.mathematik.uni-kl.de/~keilen>

PLANAR CONVEX BODIES, FOURIER TRANSFORM, LATTICE POINTS, AND IRREGULARITIES OF DISTRIBUTION

L. BRANDOLINI, A. IOSEVICH, AND G. TRAVAGLINI

ABSTRACT. Let B be a convex body in the plane. The purpose of this paper is a systematic study of the geometric properties of the boundary of B , and the consequences of these properties for the distribution of lattice points in rotated and translated copies of ρB (ρ being a large positive number), irregularities of distribution, and the spherical average decay of the Fourier transform of the characteristic function of B . The analysis makes use of two notions of “dimension” of a convex set. The first notion is defined in terms of the number of sides required to approximate a convex set by a polygon up to a certain degree of accuracy. The second is the fractal dimension of the image of the Gauss map of B . The results stated in terms of these quantities are essentially sharp and lead to a nearly complete description of the problems in question.

1. INTRODUCTION

Suppose $B \subset \mathbb{R}^2$ is a convex body: a convex compact set with nonempty interior. Many classical problems in analysis, geometry, and number theory are stated in terms of basic properties of such sets. For example, we may consider the difference between the number of lattice points inside the dilated set ρB and its area, i.e., the *discrepancy*

$$D_\rho(B) = \text{card}(\rho B \cap \mathbb{Z}^2) - \rho^2 |B|,$$

where $|\cdot|$ denotes the area. Among the many natural questions we can ask about this problem (see the section on lattice points below) is, how does the geometry of B affect the growth rate of the discrepancy function? As we shall see, there are results that do not distinguish among various convex sets. However, we shall also see that the behavior of the above discrepancy functions corresponding to different convex sets may vary dramatically, and that this behavior may be described in terms of natural and readily computable geometric quantities.

The above question on lattice points has a consequence in the study of irregularities of distribution. Suppose $\mathcal{P} = \{z_j\}_{j=1}^N$ is a distribution of N points in the unit square $U = [0, 1]^2$ treated as the torus \mathbb{T}^2 . Let B be a convex body in U with

Received by the editors February 11, 2002.

2000 *Mathematics Subject Classification.* Primary 42B10; Secondary 52A10.

Key words and phrases. Decay of Fourier transforms, convex bodies, Minkowski dimension, lattice points, irregularities of distribution.

The first and third authors are supported by MURST. The second author is supported by NSF grant DMS00-87339.

diameter smaller than 1. Assume $\varepsilon \leq 1$, $t \in \mathbb{T}^2$. Then certain sharp upper estimates for the discrepancy

$$D(\mathcal{P}, \varepsilon, t) = \sum_{j=1}^N \chi_{\varepsilon B-t}(z_j) - N \varepsilon^2 |B|$$

can be obtained from related estimates for lattice points (by a suitable trick we shall reduce to the case when N is a square, which in turn is an easy corollary).

At the heart of the lattice point and the irregularities of distribution problems is the Fourier transform of the characteristic function of B . Our approach is to study the effect of the geometric properties of B on the decay rate of the Fourier transform of the characteristic function of B and its variants. We shall then use this analysis to obtain precise information about the discrepancy functions described above.

How should we distinguish among the various convex planar sets? The lattice point problem suggests one natural approach. It was observed by Gauss that $D_\rho(B) \lesssim \rho$, since the boundary of B is one-dimensional. Consider the case when B is a unit square with sides parallel to the axes. When ρ is an integer, the boundary of ρB contains $\approx \rho$ integer lattice points, thus showing that this estimate cannot be improved. However, if B is a disc, the boundary of ρB “curves away” from the integer lattice. In fact, it is known (see [16]) that the estimate for $D_\rho(B)$ in this case is much better. These two examples suggest that the curvature of the boundary may be the key distinguishing factor among convex sets. The boundary of the square has no curvature, which leads to a poor discrepancy estimate, whereas the boundary of the disc has everywhere non-vanishing curvature, and the estimate for the discrepancy function is considerably better.

The notion of curvature alluded to in the previous paragraph is the standard geometric, or Gaussian, curvature, defined as the determinant of the differential of the Gauss map that maps each point on the boundary of a convex set to the unit normal at that point. It turns out that the geometric curvature alone does not capture the relevant properties of convex planar sets fully. To see this, let us return to the case of the unit square. While it is true that the discrepancy function is terrible if the sides of the square are parallel to the axes, the discrepancy function becomes practically non-existent, even better than the discrepancy function for the disc, if the square is rotated by a sufficiently irrational angle (see [14]). In fact, it is precisely the “flatness” of a square that keeps its boundary from hitting hardly any lattice points when it is rotated. This suggests that for “most” rotations, convex sets with “flat” boundaries behave better as far as discrepancy functions are concerned.

In this paper we consider the rotated and translated copies $\sigma^{-1}(\rho B) - t$ (where $\sigma \in SO(2)$, $t \in \mathbb{T}^2$) of the dilated body ρB (here ρ is a large positive number), and we study the L^1 mean

$$\int_{\mathbb{T}^2} \int_{SO(2)} |D_\rho(\sigma^{-1}(B) - t)| \, d\sigma$$

of the discrepancy

$$D_\rho(\sigma^{-1}(B) - t) = \text{card}((\rho\sigma^{-1}(B) - t) \cap \mathbb{Z}^2) - \rho^2 |B|.$$

The reason for choosing the L^1 mean among other L^p means will be clear soon. Let us also say that in many cases, averaging makes a discrepancy problem easier. For example, the Gauss circle problem is a basic and unsolved problem, while one can obtain (see e.g. [15] or [8]) a sharp result averaging in L^2 over translations of the discs and using only Parseval's identity and some properties of Bessel functions.

Let us go back to the geometry of B . The above observations can be exploited in a number of ways. If "flatness" is good, then the family of rotated copies of B , is better if B is close to being a polygon. This means that B is good if it can be approximated by a polygon with relatively few sides (the construction we are going to describe has been studied in [19] and [23], see also [26]). We choose an arbitrary point on the boundary of B and draw a chord to another point on the boundary of B in such a way that the maximum distance from the chord to the boundary of B is ρ^{-1} . Roughly speaking, if the number of sides of the above inscribed polygon is $\lesssim \rho^\alpha$, we say that the dimension of B is at least α (we shall explain later why for most of the paper we prefer not to consider the infimum of the α 's). Note that B is a polygon if and only if we can choose $\alpha = 0$, and if B is a circle, then $\alpha = 1/2$ works.

We can also take the following "dual" point of view. If B is close to a polygon, then its boundary ∂B has relatively few normals. A more precise way of saying this is that the area of the δ -neighborhood of the image of ∂B under the Gauss map is $\lesssim \delta^{1-d}$. If B is a disc, we can only take $d = 1$. On the other hand, we can choose $d = 0$ if and only if B is a polygon. As another example, let B be a polygon with infinitely many sides, the normals of which have apertures in the sequence $n^{-\beta}$, $\beta > 0$; it is easy to see that in this case we can take $d = (1 + \beta)^{-1}$.

Introducing the infima α^* and d^* (note that d^* is the upper Minkowski dimension of the image of the Gauss map), we have $\alpha^* \leq d^*/(d^* + 1)$, and we can also prove that this bound is the best possible. On the other hand, we can show that α^* can be as close to 0 as we want, even when d^* is away from 0.

This paper is structured as follows. We shall first describe the main analytic idea, the effect of the geometry of a convex set on the average decay of the Fourier transform of the characteristic function of B . We shall also prove that polygons provide the fastest possible decay. We shall then apply our estimates to the distribution of lattice points in convex domains and the problem of irregularities of distribution.

In this paper most of the ideas used to prove the results on the average decay are new, while almost all the applications to lattice points and irregularities of distribution are straightforward.

We conclude the introduction by noting that a notion of a dimension of a convex set may be applicable and natural in a number of interesting problems in analysis and combinatorics. For example, the Falconer distance conjecture says that if the Hausdorff dimension of a planar set is greater than 1, then the set of Euclidean distances among the points of this set has positive Lebesgue measure. However, if the Euclidean distance is replaced by the "taxi-cab" (l^1) metric, the conjecture is clearly false, and in fact the set is required to have Hausdorff dimension 2 before the same conclusion on the distance set is possible. It is reasonable to ask whether distances induced by convex sets with "intermediate dimension" provide examples of intermediate behavior in the Falconer distance problem. We hope to address this and other issues of this type in a subsequent paper.

1.1. L^p average decay of the Fourier transform. The study of the decay of the Fourier transform

$$\widehat{\chi}_B(\xi) = \int_B e^{-2\pi i \xi \cdot x} dx$$

as $|\xi| \rightarrow \infty$ is a classical subject. When ∂B has strictly positive curvature, then $|\widehat{\chi}_B(\xi)| \lesssim |\xi|^{-3/2}$. However, when ∂B contains points where the Gaussian curvature vanishes, then the above inequality is no longer true. For example, when B is a polygon and $\Theta = (\cos \theta, \sin \theta)$, then $\widehat{\chi}_B(\rho\Theta)$ decays as ρ^{-1} in some directions and as ρ^{-2} in most directions. In such cases it is useful to study the L^p spherical average decay of $\widehat{\chi}_B$, given by

$$(1.1) \quad \|\widehat{\chi}_B(\rho \cdot)\|_{L^p(\Sigma_1)},$$

where Σ_1 is the unit circle and $1 \leq p \leq \infty$. Here a basic result is Podkorytov's theorem

$$(1.2) \quad \|\widehat{\chi}_B(\rho \cdot)\|_{L^2(\Sigma_1)} \lesssim \rho^{-3/2}$$

(see [19]), where no regularity assumption on the boundary ∂B is required.

Throughout this paper $X \lesssim Y$ will mean that $X \leq cY$, with c depending only on the body B under consideration. Moreover, we shall always assume $\rho \geq 2$.

The study of (1.1) turns out to have applications to several problems, such as the distribution of lattice points in large convex domains ([20], [25], [7], [8]), irregularities of distribution ([17], [7]), summation of multiple Fourier expansions ([9], [5], [6]), and estimates for generalized Radon transforms ([21]).

The paper [8] contains the following rather complete study of (1.1) under the additional assumption that ∂B is piecewise smooth. When $p = 2$, (1.2) says that the rate of decay of (1.1) is independent of the shape of B . When $2 < p \leq \infty$, any order of decay between that of the disc and of a polygon is possible. On the other hand, when $1 \leq p < 2$, a convex body with piecewise smooth boundary behaves either like a disc or like a polygon. In particular, when P is a polygon we have the sharp bound

$$(1.3) \quad \|\widehat{\chi}_P(\rho \cdot)\|_{L^1(\Sigma_1)} \lesssim \rho^{-2} \log \rho,$$

and when B has piecewise smooth boundary, but is not a polygon, we have the sharp bound

$$(1.4) \quad \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} \lesssim c \rho^{-3/2}.$$

Actually, (1.4) is sharp whenever ∂B contains at least one point where the Gaussian curvature exists and is different from zero.

The above dichotomy, pointed out in [8], is no longer valid for arbitrary convex bodies. The existence of "chaotic" decays has been pointed out in [8, p. 553], using an abstract argument on convex sets. Unfortunately, that argument is not constructive, nor does it provide nontrivial explicit bounds for the average decay.

The main analytic tool of this paper is the L^p average decay for arbitrary convex planar bodies when $1 \leq p \leq 2$. In essence, we shall consider the L^1 average decay and the L^2 average decay. The results for intermediate exponents can be essentially obtained by interpolation. Roughly speaking, the L^2 average decay is an "all cats are grey in the dark" phenomenon, where the decay does not distinguish among the different convex bodies. On the other hand, the L^1 average decay determines, in a sense, how close a convex set is to a polygon.

1.2. Inscribed polygons. We introduce the following notation. For any $\Theta = (\cos \theta, \sin \theta)$ and any small $\delta > 0$, let

$$(1.5) \quad K_\theta = \max_{x \in B} x \cdot \Theta, \\ r(B, \delta, \theta) = \{y \in B : y \cdot \Theta = K_\theta - \delta\}.$$

We say that the chord $r(B, \delta, \theta)$ is of height δ , and we use it to define the following inscribed polygon (see also [19] or [23]).

Definition 1. Let B be a convex planar body. Choose any chord of height δ , and name it ch_1 . Move counterclockwise constructing a finite sequence of consecutive chords of height δ until you reach ch_1 . Then, if necessary, replace the last chord by one consecutive to ch_1 (hence of height not greater than δ). In this way we get a polygon inscribed in B , and we denote it by P_δ^B . Of course P_δ^B depends on the choice of ch_1 , and we should write $P_\delta^B(ch_1)$; however, none of our results depends on ch_1 , and, by a small abuse, we shall always speak about “the” inscribed polygon P_δ^B . We denote by M_δ^B be the number of sides of P_δ^B .

It has been proved in [23] that $M_\delta^B \lesssim \delta^{-1/2}$. Our first result is the following.

Theorem 2. *Let B be a convex planar body and assume $M_{\rho^{-1}}^B \lesssim \rho^\alpha$ (where $0 < \alpha < 1/2$, the cases $\alpha = 0$ and $\alpha = 1/2$ being covered by (1.3) and (1.2) respectively). Then*

$$(1.6) \quad \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} \lesssim \rho^{\alpha-2} \log \rho.$$

Moreover, for any $0 < \alpha < 1/2$, there exists a convex planar body B such that $M_{\rho^{-1}}^B \lesssim \rho^\alpha$ and, for any $\varepsilon > 0$,

$$\limsup_{\rho \rightarrow +\infty} \rho^{-\alpha+2+\varepsilon} \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} > 0.$$

All the proofs will be given in the last section of the paper (§3).

Before going on, we want to discuss the above theorem. The first step in the proof is to show that

$$\int_0^{2\pi} |\widehat{\chi}_B(\rho\Theta)| \, d\theta \lesssim \int_0^{2\pi} |\widehat{\chi}_{P_{\rho^{-1}}^B}(\rho\Theta)| \, d\theta$$

(see Definition 1). We are therefore reduced to estimating the average decay for a polygon with $\lesssim \rho^\alpha$ sides. The second step simply consists in recalling that the implicit constant in (1.3) depends on the number of sides of the polygon P , and that after reading the proofs in [7] or [8] one can rewrite (1.3) in the following way:

$$(1.7) \quad \int_0^{2\pi} |\widehat{\chi}_P(\rho\Theta)| \, d\theta \leq cN\rho^{-2} \log \rho,$$

where N is the number of sides of the polygon P , and the constant c is absolute (there is no loss of generality in assuming that the length of the boundary ∂P is ≤ 1). Putting ρ^α in place of N , we then get (1.6).

At this point one should expect to have gotten a *poor* result using the trivial estimate (1.7). The counterexample in the theorem shows that this is not so.

1.3. The image of the Gauss map. At every point of ∂B there is a left and a right tangent, therefore a left $(-)$ and a right $(+)$ outward normal. Let $\pi^\pm : \partial B \rightarrow \Sigma_1$ be the map sending each point in ∂B to the left/right normal. Also let

$$(1.8) \quad \Delta^B = \pi^-(\partial B) \cup \pi^+(\partial B).$$

We identify Σ_1 with the interval $[0, 2\pi)$. For every $\theta \in [0, 2\pi)$ we denote by $d(\theta, \Delta^B)$ the distance between θ and Δ^B . For a given small δ , let

$$(1.9) \quad \Delta_\delta^B = \{x \in [0, 2\pi) : d(x, \Delta^B) < \delta\}$$

be the δ -neighborhood of Δ^B .

Theorem 3. *Let $0 < d < 1$. Assume*

$$(1.10) \quad |\Delta_\delta^B| \lesssim \delta^{1-d}.$$

Then

$$(1.11) \quad \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} \lesssim \rho^{\frac{d}{d+1}-2}.$$

Moreover, there exists a convex body B satisfying $|\Delta_\delta^B| \lesssim \delta^{1-d}$ and such that

$$\limsup_{\rho \rightarrow +\infty} \rho^{-\frac{d}{d+1}+2+\varepsilon} \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} > 0$$

for any $\varepsilon > 0$.

The proof will be given in the last section.

Remark 4. Again, the cases $d = 0$ and $d = 1$ are covered by (1.3) and (1.2) respectively.

Remark 5. We point out that the infimum of the numbers d such that $|\Delta_\delta^B| \lesssim \delta^{1-d}$ is just the upper Minkowski dimension of Δ^B , that is, the number

$$d^* = \limsup_{\delta \rightarrow 0} \left(\log_{1/\delta} (|\Delta_\delta^B|/\delta) \right).$$

It is therefore possible to restate Theorem 3 in a form such as “If $d > d^*$, then (1.11) holds”. However, we prefer to keep the original statement in Theorem 3 for the following two reasons. First, the L.H.S. in (1.10) is the quantity that actually arises in the proof. Second, we do not want to confuse naturally different objects, such as the polygons with finitely many sides and certain polygons with infinitely many sides (e.g. with an exponentially decreasing sequence of slopes) which share $d^* = 0$ with the polygons with finitely many sides. For similar reasons we did not introduce the infimum α^* of the α ’s in Theorem 2. On the contrary, we shall introduce α^* and d^* in the following section in order to get a neater comparison.

1.4. Comparing the previous arguments. For any B we denote by d^* the Minkowski dimension of Δ^B (see the above remark). We also denote by α^* the infimum of the α ’s such that $M_{\rho^{-1}}^B \leq c_\alpha \rho^\alpha$. We have the following theorem.

Theorem 6. *Let B be a convex planar body. Then*

$$\alpha^* \leq \frac{d^*}{d^* + 1}.$$

Moreover, there exists B for which equality holds.

The proof will be given in the last section.

Remark 7. Theorem 6 exhibits an upper bound for α^* in terms of d^* . A lower bound in terms of d^* does not exist in general, since we can construct a family of convex bodies with the same $d^* > 0$ but α^* arbitrarily close to 0.

The proof will be given in the last section.

The situation is different if we add geometric assumptions on B .

Theorem 8. *Suppose B is inscribed in a disc (i.e., B is the convex hull of a subset of a circle). Then $\alpha^* = d^*/2$.*

The proof will be given in the last section.

The circle in the previous statement can be replaced by a closed convex smooth curve with everywhere positive Gaussian curvature.

Remark 9. By appealing to Theorem 2 and Theorem 6 we immediately get the following inequality, which is slightly weaker than the one in Theorem 3:

$$\|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} \lesssim \rho^{\frac{d}{d+1}-2+\varepsilon}.$$

1.5. A lower bound for all convex bodies. The main results in this paper deal with “intermediate” cases between polygons and convex bodies having a smooth convex arc in the boundary. These cases turn out to be extreme. Indeed, Podkorytov’s theorem is a uniform (with respect to the choice of B) upper bound, while the following theorem gives a uniform lower bound for the L^1 average decay of the Fourier transform.

Theorem 10. *Let B be a convex body in \mathbb{R}^2 . Then*

$$\limsup_{\rho \rightarrow +\infty} \rho^2 \log^{-1} \rho \|\widehat{\chi}_B(\rho \cdot)\|_{L^1(\Sigma_1)} > 0.$$

The proof will be given in the last section.

2. APPLICATIONS

2.1. Lattice points. Let B be a planar convex body, and let $\sigma \in SO(2)$ and $t \in \mathbb{T}^2$. We consider the discrepancy

$$(2.1) \quad D_\rho(B) = \text{card}(\rho B \cap \mathbb{Z}^2) - \rho^2 |B|,$$

where $|\cdot|$ denotes the area. The results in the previous section and some arguments in [20], [25], [7], and [8] allow us to obtain several upper and lower bounds for averages of the discrepancy (2.1) over rotations or rotations and translations. As a first example, it has been proved in [15], [25], and [7] that, for a polygon P , (1.3) implies

$$\int_{SO(2)} |D_\rho(\sigma^{-1}(P))| \, d\sigma \lesssim \log^2 \rho.$$

As another example, one can use (1.2) to show that for any convex planar body B ,

$$(2.2) \quad \left\{ \int_{\mathbb{T}^2} \int_{SO(2)} |D_\rho(\sigma^{-1}(P) - t)|^2 \, d\sigma dt \right\}^{1/2} \lesssim \rho^{1/2}.$$

(See e.g. [15] or [8].) Note that (2.2) is false without the integration in t , as the case of a disc and Hardy’s Ω -result (see [16]) show.

Again we focus on the case $p = 1$, and we have the following result, which follows easily from Theorem 2 and some known arguments (see e.g. [15], [25] or [7]).

Theorem 11. *Let B be a planar convex body such that $M_{\rho^{-1}}^B \lesssim \rho^\alpha$ ($0 < \alpha < 1/2$). Then*

$$(2.3) \qquad \int_{\mathbb{T}^2} \int_{SO(2)} |D_\rho(\sigma^{-1}(B) - t)| \, d\sigma dt \lesssim \rho^{\frac{2\alpha}{2\alpha+1}} \log \rho.$$

Moreover, for every such α there exists a body B satisfying

$$\limsup_{\rho \rightarrow +\infty} \rho^{-\alpha+\varepsilon} \int_{\mathbb{T}^2} \int_{SO(2)} |D_\rho(\sigma^{-1}(B) - t)| \, d\sigma dt > 0,$$

for any $\varepsilon > 0$.

The proof will be given in the last section.

Remark 12. The cases $\alpha = 0$ and $\alpha = 1/2$ are known; see e.g. [7] and [8] respectively.

2.2. Irregularities of distribution. Suppose $\mathcal{P} = \{z_j\}_{j=1}^N$ is a distribution of N points in the unit square $U = [0, 1]^2$ treated as the torus \mathbb{T}^2 . Let B be a convex body in U with diameter smaller than 1. Assume $\varepsilon \leq 1$, $\sigma \in SO(2)$, and $t \in \mathbb{T}^2$. The study of the discrepancy

$$D(\mathcal{P}, \varepsilon, \sigma, t) = \sum_{j=1}^N \chi_{\varepsilon \sigma^{-1} B - t}(z_j) - N \varepsilon^2 |B|$$

has a long history (see e.g. the references in [2] and [17, ch. 6]). A typical result is the following theorem, due to Beck [1] and Montgomery [17, ch. 6] (see also [7]).

Theorem 13. *Let B be a convex body in $U = [0, 1]^2$ with diameter smaller than 1. Then, for every distribution $\mathcal{P} = \{z_j\}_{j=1}^N$ in U ,*

$$\left\{ \int_0^1 \int_{SO(2)} \int_{\mathbb{T}^2} |D(\mathcal{P}, \varepsilon, \sigma, t)|^2 \, dt \, d\sigma \, d\varepsilon \right\}^{1/2} \gtrsim N^{1/4}.$$

The above result is sharp, since Beck and Chen [3] proved the following upper bound.

Theorem 14. *Let B be a convex body in $U = [0, 1]^2$ with diameter smaller than 1. Then for every positive integer N there exists a distribution \mathcal{P} of N points such that*

$$(2.4) \qquad \left\{ \int_0^1 \int_{SO(2)} \int_{\mathbb{T}^2} |D(\mathcal{P}, \varepsilon, \sigma, t)|^2 \, dt \, d\sigma \, d\varepsilon \right\}^{1/2} \lesssim N^{1/4}.$$

The above upper bound can be improved after replacing the L^2 norm with the L^1 norm. Indeed, Beck and Chen [4] proved the following result.

Theorem 15. *Let P be a convex polygon in $U = [0, 1]^2$ with diameter smaller than 1. Then for every positive integer N there exists a distribution \mathcal{P} of N points such that*

$$(2.5) \qquad \int_0^1 \int_{SO(2)} \int_{\mathbb{T}^2} |D(\mathcal{P}, \varepsilon, \sigma, t)| \, dt \, d\sigma \, d\varepsilon \lesssim \log^2 N.$$

The next result follows easily from Theorem 11, [7] and [8]. The case $\alpha = 0$ provides a different proof of (2.5). In the same way one can get a different proof of the L^2 result in (2.4) too. We point out that appealing to lattice point results does not work for L^p norms when $p > 2$ and the body is a polygon (see [11]).

Theorem 16. *Let B be a convex body in $U = [0, 1]^2$ with diameter smaller than 1 and such that $M_{\rho^{-1}}^B \lesssim \rho^\alpha$. Then for every positive integer N there exists a distribution \mathcal{P} of N points satisfying*

$$\int_{\mathbb{T}^2} \int_{SO(2)} |D(\mathcal{P}, \sigma, t)| \, d\sigma dt \lesssim \begin{cases} \log^2 N & \text{when } \alpha = 0, \\ N^{\frac{\alpha}{1+2\alpha}} \log N & \text{when } 0 < \alpha < 1/2, \\ N^{1/4} & \text{when } \alpha = 1/2, \end{cases}$$

where $D(\mathcal{P}, \sigma, t) = D(\mathcal{P}, 1, \sigma, t)$.

The proof will be given in the last section.

3. PROOFS

The following known result (see e.g. [10], [19], [8]) will be used throughout the paper.

Lemma 17. *Let B be a convex body in \mathbb{R}^2 . Using the notation in (1.5), we have*

$$|\widehat{\chi}_B(\rho\Theta)| \lesssim \rho^{-1} [|r(B, \rho^{-1}, \theta)| + |r(B, \rho^{-1}, \theta + \pi)|],$$

where $|\cdot|$ denotes the length of the chord.

We define

$$\widetilde{d}(\theta, \Delta^B) = \min(d(\theta, \Delta^B), d(\theta + \pi, \Delta^B)),$$

and we deduce the following lemma.

Lemma 18. *For every $\theta \notin \Delta^B$ we have*

$$|\widehat{\chi}_B(\rho\Theta)| \lesssim \frac{1}{\rho^2 \widetilde{d}(\theta, \Delta^B)}.$$

Proof. Let $\theta \notin \Delta^B$ (say $\theta = -\pi/2$). Assume that ∂B passes through the origin and that B lies in the upper half-plane. It follows that in a neighborhood of the origin ∂B is the graph of a nonnegative convex function, say $y = \varphi(x)$, satisfying $\varphi(0) = 0$ and $\varphi'(0-) < 0 < \varphi'(0+)$, where $\varphi'(0-)$ and $\varphi'(0+)$ denote the left and the right derivative at the origin respectively. Let

$$E = \{(x, y) \in \mathbb{R}^2 : y > \varphi'(0-)x \text{ and } y > \varphi'(0+)x\}.$$

By convexity, $B \subset E$, and therefore

$$|r(B, \rho^{-1}, \theta)| \leq \frac{1}{\rho \varphi'(0+)} + \frac{1}{\rho |\varphi'(0-)|} \leq \frac{2}{\rho \min(\varphi'(0+), |\varphi'(0-)|)}.$$

To complete the proof it is enough to observe that

$$\min(\varphi'(0+), |\varphi'(0-)|) \approx d(\theta, \Delta^B)$$

and to apply the previous lemma. □

The following lemmas will be needed in the proof of Theorem 2.

Lemma 19. *Let $R \geq 1$ and $0 < \beta < \pi/4$. Assume $R\beta < 1/2$. Denote by $C = C(\beta, R)$ the convex hull of the set*

$$\{R \exp(i\theta) : -\beta \leq \theta \leq \beta\} \cup \{P\},$$

where the point P has distance 1 from the points $Re^{\pm i\beta}$ and satisfies $|P| \leq R$. Then there exist positive constants c_1 and c_2 such that if $R\rho\beta^2 \geq c_1$ then we have

$$|\widehat{\chi}_C(\rho\Theta)| \geq c_2 R^{1/2} \rho^{-3/2}$$

for every $|\theta| \leq \beta/2$.

Proof. Integrating by parts, we are reduced to estimating

$$(3.1) \quad \rho^{-1} \int_{\partial C} n(x) \cdot \Theta \exp(2\pi i \rho \Theta \cdot x) dx.$$

The boundary ∂C consists of two segments and an arc. In order to control the latter we reduce to the oscillatory integral

$$\left| \int_{-R\beta}^{R\beta} \exp\left(i\rho \frac{t^2}{R}\right) dt \right| = \left| R\beta \int_{-1}^1 \exp(i\rho R\beta^2 u) du \right| \geq c R^{1/2} \rho^{-1/2}$$

for $\rho R\beta^2$ large enough. The two segments have length 1, and their contribution in (3.1) is $O(\rho^{-2})$. \square

Lemma 20. *Let $R > 1$ and $0 < \beta < \pi/4$. Assume $R\beta < \frac{1}{2}$. For any $N \geq 1$, let $B = B(\beta, R, N)$ be the convex hull of the set*

$$\{R \exp(2\pi i k \beta / N), k = -N, \dots, N\} \cup \{P\},$$

where, as before, the point P has distance 1 from the points $Re^{\pm i\beta}$ and satisfies $|P| \leq R$. Then there exist absolute constants c_1 , c_2 , and c_3 such that whenever $\rho \geq 2$ and

$$(3.2) \quad \frac{c_1}{\beta^2} \leq R\rho \leq \frac{c_2}{\beta^2} \frac{N^2}{\log^2 N},$$

we have, for any $-\beta/2 \leq \theta \leq \beta/2$,

$$|\widehat{\chi}_B(\rho\Theta)| \geq c_3 R^{1/2} \rho^{-3/2}.$$

Proof. Let $C = C(\beta, R)$ be as in Lemma 19. By (3.2) and Lemma 19 we have

$$|\widehat{\chi}_C(\rho\Theta)| \geq c R^{1/2} \rho^{-3/2}$$

when $-\beta/2 \leq \theta \leq \beta/2$.

We now study the Fourier transform $\widehat{\chi}_{C \setminus B}$. We claim that

$$(3.3) \quad |\widehat{\chi}_{C \setminus B}(\rho\Theta)| \leq c\beta\rho^{-1} \frac{\log N}{N} R$$

uniformly in θ . Indeed, $C \setminus B$ is the union of $2N$ "lunes" ℓ_1, \dots, ℓ_{2N} (each lune is a convex set bounded by a segment in B and by a portion of the arc in C , see Figure 1) and, for any θ ,

$$\widehat{\chi}_{C \setminus B}(\rho\Theta) = \widehat{f}(\rho),$$

where $f = f_\theta$ is defined by

$$\begin{aligned} f(s) &= |C \setminus B \cap \{\xi \in \mathbb{R}^2 : \xi \cdot \Theta = s\}| \\ &= \sum_{k=1}^{2N} |\ell_k \cap \{\xi \in \mathbb{R}^2 : \xi \cdot \Theta = s\}| \\ &= \sum_{k=1}^{2N} f_k(s). \end{aligned}$$

Note that, for any given s , the above sum contains at most two terms. It is enough to consider one of them, i.e., we assume $0 \leq \theta \leq \pi$. Moreover, we are reduced to studying the case $0 \leq \theta < \beta/N$, the other cases being similar. In order to bound $\hat{f}(\rho)$, we estimate the total variation V_f of the function $f(s)$, which is the length of the vertical segment in the k th lune. Now observe that

$$V_{f_k} \leq c\beta N^{-1}k^{-1}R$$

whenever $k \geq 1$ (see Figure 1).

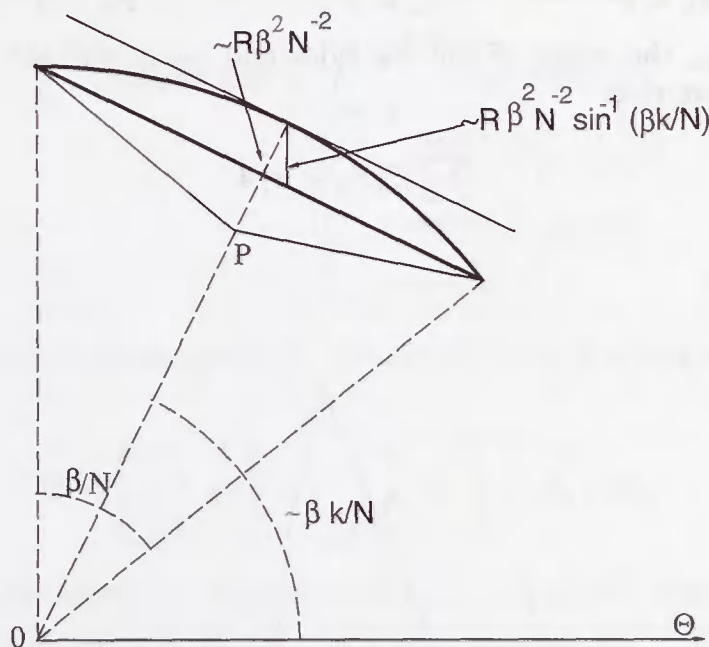


FIGURE 1

Summing on k (there are N terms when $\theta = 0$ and $N + 1$ terms when $0 < \theta < \beta/N$), we get (3.3).

Finally, for suitable choices of c_1 and c_2 in (3.2) we get

$$\begin{aligned} |\hat{\chi}_B(\rho\Theta)| &\geq |\hat{\chi}_C(\rho\Theta)| - |\hat{\chi}_{B \setminus C}(\rho\Theta)| \\ &\geq c_3 R^{1/2} \rho^{-3/2} - c_4 \beta \rho^{-1} \frac{\log N}{N} R \\ &\geq c_5 \rho^{-3/2} R^{1/2}. \end{aligned}$$

□

Proof of Theorem 2. We start with the upper bounds in (1.6). Let $P_{\rho^{-1}}^B$ be as in Definition 1. Let $\tilde{P}_{\rho^{-1}}^B$ be the smallest polygon having sides parallel to those of

P_ρ^B and containing B . It is not difficult to see that for ρ sufficiently large,

$$|r(B, \rho^{-1}, \theta)| \lesssim \left| r(\tilde{P}_\rho, c\rho^{-1}, \theta) \right|,$$

where again the implicit constant depends only on B . By Lemma 17 we have

$$\begin{aligned} |\widehat{\chi}_B(\rho\Theta)| &\lesssim \rho^{-1} |r(B, \rho^{-1}, \theta)| \\ &\lesssim \rho^{-1} \left| r(\tilde{P}_{\rho^{-1}}, c\rho^{-1}, \theta) \right|. \end{aligned}$$

Hence, by the proof of (1.7) in [7] or [8],

$$\rho^{-1} \int_0^{2\pi} \left| r(\tilde{P}_{\rho^{-1}}, c\rho^{-1}, \theta) \right| d\theta \leq cM_{\rho^{-1}}^B \rho^{-2} \log(\rho) \leq c\rho^{-2+\alpha} \log(\rho),$$

thereby proving (1.6).

We now show that (1.6) is essentially sharp. Let $B = B(\beta, R, N)$ be as in Lemma 20 and consider the sets $B_h = B(\beta_h, R_h, N_h)$, $h = 1, 2, 3, \dots$, where, for any small $\varepsilon > 0$,

$$R_h = 2^{(1-2\alpha)h}, \quad \beta_h = 2^{h(2\alpha-1-\varepsilon)}, \quad N_h = 2^{h\alpha}.$$

We denote by γ_h the union of the N_h sides and by ζ_h the arc where they are inscribed. Observe that

$$(3.4) \qquad \sum_{h=n_0}^{+\infty} \beta_h R_h < \pi/4$$

for a suitable n_0 .

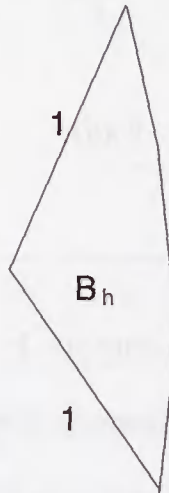


FIGURE 2

We recall that each B_h has the shape in Figure 2, i.e., it is a convex polygon consisting of two sides of length 1 and of N_h sides coming from a regular polygon of large radius R_h . Let E_h be the rotated and translated copy of every B_h , so that, moving counterclockwise, $E_{n_0} = B_{n_0}$ and two consecutive E_h 's have disjoint

interior and share a side (of length 1), while the union of the arcs ζ_h is a convex piecewise smooth curve. We write

$$(3.5) \quad B = \left(\bigcup_{j=n_0}^{h-1} E_j \right) \cup E_h \cup \left(\bigcup_{j=h+1}^{\infty} E_j \right) = \tilde{E}_h \cup E_h \cup E_h^{\#}.$$

By the condition (3.4), B is a convex set. Now let $\rho_h = 2^h$. Let $p_h = \sum_{j=n_0}^h \beta_j$. Since (3.2) is satisfied, Lemma 20 implies

$$|\widehat{\chi}_{E_h}(\rho_h \Theta)| \geq c R_h^{1/2} \rho_h^{-3/2} = c 2^{-h(\alpha+1)}$$

for

$$(3.6) \quad p_h + \frac{1}{3}\beta_h < \theta < p_h + \frac{2}{3}\beta_h.$$

We then estimate the contribution of the convex sets \tilde{E}_h and $E_h^{\#}$, using Lemma 18. Indeed, since θ satisfies (3.6), we obtain, for any h ,

$$\left| \widehat{\chi}_{\tilde{E}_h}(\rho_h \Theta) \right| + \left| \widehat{\chi}_{E_h^{\#}}(\rho_h \Theta) \right| \leq c \beta_h^{-1} \rho_h^{-2}.$$

We then have

$$\begin{aligned} \int_0^{2\pi} |\widehat{\chi}_B(\rho_h \Theta)| d\theta &\geq \int_{p_h + \frac{1}{3}\beta_h}^{p_h + \frac{2}{3}\beta_h} |\widehat{\chi}_B(\rho_h \Theta)| d\theta \\ &\geq \left| c_1 \beta_h R_h^{1/2} \rho_h^{-3/2} - c_2 \rho_h^{-2} \right| \\ &\geq \left| c_1 2^{h(\alpha-\varepsilon-2)} - c_2 2^{-2h} \right| \\ &\geq c_3 \rho_h^{-2+\alpha-\varepsilon}. \end{aligned}$$

To complete the proof we estimate $M_{\rho^{-1}}^B$. Given $\rho \geq 2$, let H satisfy $2^H \leq \rho < 2^{H+1}$. Here we split

$$(3.7) \quad B = \left(\bigcup_{j=n_0}^H E_j \right) \cup \left(\bigcup_{j=H+1}^{+\infty} E_j \right) = B_a \cup B_b.$$

Observe that the first term is a polygon with $\sum_{j=n_0}^H N_j \lesssim 2^{H\alpha}$ sides. Now consider that for any convex polygon Q and any δ , the number M_{δ}^Q cannot exceed the number of sides of Q . Therefore the contribution of B_a to $M_{\rho^{-1}}^B$ is $\lesssim 2^{H\alpha} = \rho^{\alpha}$. As for B_b , we note that the length of $\bigcup_{j=H+1}^{+\infty} \zeta_j$ is comparable to the length of ζ_H , while the chords of height ρ^{-1} are longer, since $\bigcup_{j=H+1}^{+\infty} \zeta_j$ comes from flatter arcs. Therefore there are fewer chords than for ζ_H . We have therefore proved that $M_{\rho^{-1}}^B \lesssim \rho^{\alpha}$. \square

Proof of Theorem 3. Let $\Omega_{\rho} = \Delta_{\rho^{-1/(d+1)}}^B$. In order to estimate

$$I(\rho) = \int_0^{2\pi} |\widehat{\chi}_B(\rho \Theta)| d\theta$$

we write

$$I(\rho) = \int_{\Omega_{\rho}} |\widehat{\chi}_B(\rho \Theta)| d\theta + \int_{[0, 2\pi] \setminus \Omega_{\rho}} |\widehat{\chi}_B(\rho \Theta)| d\theta = I_1 + I_2.$$

To estimate I_1 we use the Cauchy-Schwarz inequality, the fact that $|\Delta_\delta^B| \lesssim \delta^{1-d}$, and (1.2):

$$\begin{aligned} I_1 &\leq |\Omega_\rho|^{1/2} \left\{ \int_0^{2\pi} |\widehat{\chi}_B(\rho\Theta)|^2 d\theta \right\}^{1/2} \\ &\lesssim \rho^{(d-1)/(2d+2)} \rho^{-3/2} \\ &= c\rho^{-2+\frac{d}{d+1}}. \end{aligned}$$

In order to estimate I_2 we use Lemma 18:

$$\begin{aligned} I_2 &\lesssim \sum_{k=0}^{(d+1)^{-1} \log \rho} \int_{\Delta_{2^{-k}}^B \setminus \Delta_{2^{-k-1}}^B} \frac{c}{\rho^2 \widetilde{d}(\theta, \Delta^B)} d\theta \\ &\lesssim \rho^{-2} \sum_{k=0}^{(d+1)^{-1} \log \rho} 2^k |\Delta_{2^{-k}}^B| \\ &\lesssim \rho^{-2} \sum_{k=0}^{(d+1)^{-1} \log \rho} 2^k 2^{-k(1-d)} \\ &\lesssim \rho^{-2} \sum_{k=0}^{(d+1)^{-1} \log \rho} 2^{kd} \\ &= c\rho^{-2+\frac{d}{d+1}}. \end{aligned}$$

In order to give a counterexample we use the body B constructed in the proof of Theorem 2. Again we consider the sets $B_h = B(\beta_h, R_h, N_h)$, $h = 1, 2, \dots$, where now

$$R_h = 2^{h\frac{1-d}{1+d}}, \quad \beta_h = 2^{h(\frac{d-1}{d+1}-\varepsilon)}, \quad N_h = 2^{h\frac{d}{d+1}},$$

while $\rho_h = 2^h$. Arguing as in the proof of the previous theorem, we get, for every h ,

$$\rho_h^{2-\frac{d}{1+d}+\varepsilon} \int_0^{2\pi} |\widehat{\chi}_B(\rho_h\Theta)| d\theta \geq c.$$

To complete the proof it is enough to show that $|\Delta_\delta^B| \lesssim \delta^{1-d}$. We identify Δ_δ^B with a subset of $[0, \pi/2]$, and we observe that

$$\Delta_\delta^B \cap \left[\sum_{j \leq H-1} \beta_j, \sum_{j \leq H} \beta_j \right]$$

consists of N_H points at distance β_H/N_H . Given $\delta > 0$, we choose H so that

$$\frac{\beta_H}{N_H} \leq \delta < \frac{\beta_{H-1}}{N_{H-1}}.$$

Hence

$$\beta_H \leq \left(\frac{\beta_H}{N_H} \right)^{1-d} \approx \delta^{1-d}.$$

We now split $B = B_a \cup B_b$ as in (3.7). The contribution of B_a to $|\Delta_\delta^B|$ is

$$\delta \sum_{j \leq H} N_j \approx \delta N_H \approx \beta_H \lesssim \delta^{1-d},$$

while the contribution of B_b to $|\Delta_\delta^B|$ is bounded by

$$\sum_{j>H} \beta_j \lesssim \beta_H \lesssim \delta^{1-d}.$$

□

The next proof follows an argument in [23].

Proof of Theorem 6. Let ch_j be a side of $P_{\rho^{-1}}^B$ having endpoints x_j and y_j . Assume that, moving counterclockwise along the boundary of B , the point x_j comes before y_j . Denote by φ_j the direction of the right normal in x_j and by ψ_j the direction of the left normal in y_j . First observe that

$$(3.8) \quad |ch_j| |\varphi_j - \psi_j| \gtrsim \rho^{-1}.$$

((3.8) follows by convexity when $|\varphi_j - \psi_j| \geq \pi/4$ and by a trigonometric computation when $|\varphi_j - \psi_j| < \pi/4$.) Let $\alpha > \alpha^*$. Summing up and applying the Hölder inequality, we get

$$\begin{aligned} \rho^{-\alpha} M_{\rho^{-1}}^B &\lesssim \sum_j |ch_j|^\alpha |\varphi_j - \psi_j|^\alpha \\ &\leq \left\{ \sum_j |ch_j| \right\}^\alpha \left\{ \sum_j |\varphi_j - \psi_j|^{\frac{\alpha}{1-\alpha}} \right\}^{1-\alpha} \\ &\leq |\partial B|^\alpha \left(\sum_j |\varphi_j - \psi_j|^{\frac{\alpha}{1-\alpha}} \right)^{1-\alpha}, \end{aligned}$$

where the sum is on the $M_{\rho^{-1}}^B$ sides of the polygon $P_{\rho^{-1}}$. It remains to show that $\sum_j |\varphi_j - \psi_j|^{\frac{\alpha}{1-\alpha}}$ is bounded by a constant independent of $P_{\rho^{-1}}$. Let

$$Z_k = \{j : 2^{-k}\pi < |\varphi_j - \psi_j| \leq 2^{1-k}\pi\}.$$

Now observe that if $j \in Z_k$, then the interval $(\varphi_j, \psi_j) \subseteq \Delta_{2^{-k}\pi}^B$. Now choose d such that $d^* < d < \frac{\alpha}{1-\alpha}$. Then

$$2^{-k}\pi \operatorname{card}(Z_k) \leq |\Delta_{2^{-k}\pi}^B| \lesssim 2^{-k(1-d)},$$

so that $\operatorname{card}(Z_k) \lesssim 2^{kd}$, and therefore

$$\begin{aligned} \sum_j |\varphi_j - \psi_j|^{\frac{\alpha}{1-\alpha}} &\leq \sum_{k=0}^{+\infty} \sum_{j \in Z_k} |\varphi_j - \psi_j|^{\frac{\alpha}{1-\alpha}} \\ &\lesssim \sum_{k=0}^{+\infty} 2^{kd} 2^{-k \frac{\alpha}{1-\alpha}} \\ &= \sum_{k=0}^{+\infty} 2^{-k(\frac{\alpha}{1-\alpha} - d)} \\ &< +\infty. \end{aligned}$$

The sharpness of the inequality $\alpha^* \leq \frac{d^*}{d^*+1}$ follows from the common counterexample in the proof of Theorem 2 and Theorem 3. □

Proof of Remark 7. Let $\gamma > 1$ and $\beta > 0$. For $n \geq 1$, let $x_n = n^{-\beta}$ and $y_n = n^{-\beta\gamma}$. Let B denote the convex hull of the infinite points (x_n, y_n) . We claim that the polygon $P_{\rho^{-1}}$ associated to B satisfies

$$M_{\rho^{-1}}^B \lesssim \rho^{\frac{1}{\gamma\beta}}$$

(hence $\alpha^* \leq 1/\gamma\beta$). Indeed, choose

$$ch_1 = B \cap \left\{ (x, y) \in \mathbb{R}^2 : y = \frac{1}{\rho} \right\}$$

as the first side of $P_{\rho^{-1}}$. The number of sides of B located on the right of ch_1 is $\approx \rho^{1/\gamma\beta}$, and the claim follows since for any polygon D with finitely many sides and any ρ we have $M_{\rho^{-1}}^D \leq \#(\text{sides of } D)$. On the other hand, one checks that B satisfies

$$|\Delta_\delta^B| \lesssim \delta^{1 - \frac{1}{\beta(\gamma-1)+1}}$$

and the exponent is the best possible (i.e., $d^* = 1/(\beta(\gamma-1)+1)$).

If we now choose $\gamma = 1 + 1/\beta$, we get $d^* = 1/2$ and α^* arbitrarily small (since β can be large). \square

Proof of Theorem 8. We show that $\alpha^* = d^*/2$ whenever B is inscribed in a disc, namely when B is the convex hull of a subset of a circle.

Let $P_{\rho^{-1}}^B$ be as in Definition 1 and assume $\alpha > \alpha^*$, hence $M_{\rho^{-1}}^B \lesssim \rho^\alpha$. Let x_1, x_2, \dots be the vertices of $P_{\rho^{-1}}^B$. See Figure 3.

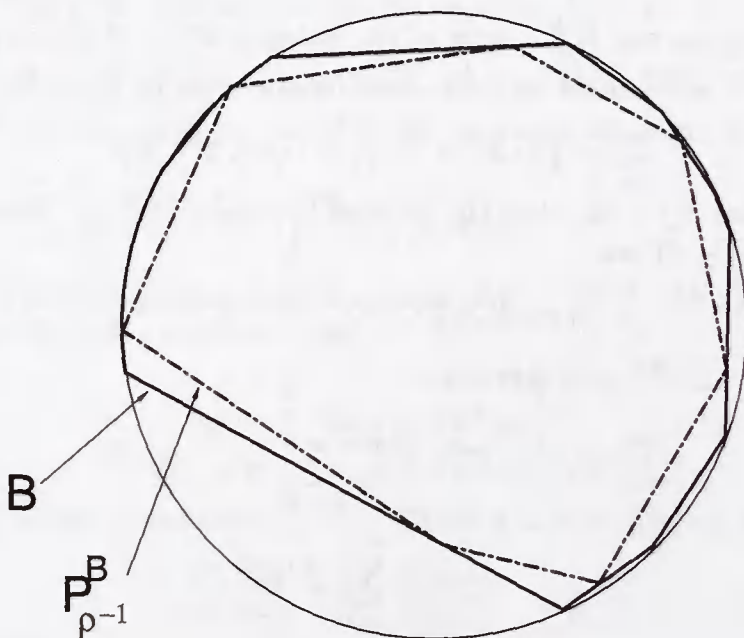


FIGURE 3

Let B_1, B_2, \dots be discs of radius $\rho^{-1/2}$ centered at the above vertices. Since B is the convex hull of a subset of a given circle C , there exists a constant c such that, for any j , we are in at least one of the following two cases:

either

i) $cB_j \cup cB_{j+1}$ contains the arc in ∂B connecting x_j and x_{j+1} ,

or

ii) the part of ∂B connecting x_j and x_{j+1} and not contained in $cB_j \cup cB_{j+1}$ is a segment.

Indeed, assume that i) and ii) fail. Then the arc in ∂B connecting x_j and x_{j+1} must touch the unit circle C outside of the discs cB_j or cB_{j+1} , at a point having distance $\approx \rho^{-1}$ from the side of $P_{\rho^{-1}}^B$ connecting x_j and x_{j+1} . Now observe that this latter can be extended to a chord of C at distance $\approx \rho^{-1}$ from ∂C . Then, for a suitable c , the discs cB_j and cB_{j+1} cannot be distinct.

The above implies that, for $\alpha > \alpha^*$,

$$\Delta_{\rho^{-1/2}}^B \subseteq c_1 \pi^\pm \left(\partial B \cap \left(\bigcup_{j=1}^{c\rho^\alpha} cB_j \right) \right),$$

and therefore

$$\left| \Delta_{\rho^{-1/2}}^B \right| \lesssim \sum_{j=1}^{c\rho^\alpha} \rho^{-1/2} \approx \rho^{\alpha-1/2} = \left(\rho^{-1/2} \right)^{1-2\alpha};$$

hence, in this case, $d^* \leq 2\alpha^*$.

We now prove that $\alpha^* \leq d^*/2$. Let $\bar{\alpha} < \alpha^*$. Then there exists a sequence $\rho_k \rightarrow +\infty$ such that $M_{\rho_k}^{B-1} \gtrsim \rho_k^{\bar{\alpha}}$. We claim that there exist $\approx \rho_k^{\bar{\alpha}}$ points in Δ^B that are $\approx \rho_k^{-1/2}$ separated. Postponing for a moment the proof of the claim, we conclude that

$$\left| \Delta_{\rho_k^{-1/2}}^B \right| \gtrsim \rho_k^{\bar{\alpha}-1/2} = \left(\rho_k^{-1/2} \right)^{1-2\bar{\alpha}},$$

which implies that the Minkowski dimension d^* of Δ^B cannot be smaller than $2\bar{\alpha}$, and therefore $d^* \geq 2\alpha^*$. \square

Proof of the claim. Let ch_j , φ_j and ψ_j be as in the proof of Theorem 6, and define

$$S_a = \left\{ j : |\varphi_j - \psi_j| > \rho_k^{-1/2} \right\},$$

$$S_b = \left\{ j : |\varphi_j - \psi_j| \leq \rho_k^{-1/2} \right\}.$$

It is enough to prove that whenever $j \in S_b$ we have $|\varphi_j - \psi_j| \gtrsim c\rho_k^{-1/2}$. Since B is inscribed in a (unit) circle, a simple geometric argument shows that if $|\varphi_j - \psi_j| \leq \rho_k^{-1/2}$, then the chord ch_j (which is a chord of B of height ρ_k^{-1}) can be continued to a chord of the circle of height $\approx \rho_k^{-1}$ and therefore of length $\approx \rho_k^{-1/2}$. It follows that $|ch_j| \lesssim \rho_k^{-1/2}$, and (3.8) yields $|\varphi_j - \psi_j| \gtrsim c\rho_k^{-1/2}$ for any $j = 1, \dots, c\rho^\alpha$. \square

The following lemma will be needed in the proof of Theorem 10. The proof depends on an easy modification of an argument in [27].

Lemma 21. *Let B be a convex planar body containing a large disc of radius r . Let g be a smooth nonnegative function supported in the set $\{t + v\}_{t \in B, |v| \leq 1}$ such that $g(t) = 1$ when $t \in B$ and $\text{dist}(t, \partial B) \geq 1$. Then there exists a constant c , independent of r , such that*

$$\|\widehat{g}\|_{L^1(\mathbb{R}^2)} \geq c \log^2 r.$$

Proof. We first need the following known inequality (see e.g. [24] or [13]). Let $h \in L^1(\mathbb{R})$ satisfy $\widehat{h} \in L^1(\mathbb{R})$ and $\widehat{h}(u) = 0$ for $u \leq 0$. Then

(3.9)

$$\int_{-\infty}^{+\infty} |h(x)| \, dx \geq c \int_1^{+\infty} \frac{1}{u} \left| \widehat{h}(u) \right| \, du.$$

A quick proof of (3.9) follows. Because of [12, p. 584] we can assume $\widehat{h}(u) \geq 0$. We then consider the odd real function s defined by $s(x) = -i(1-x)_+$ for $x > 0$, the Fourier transform of which is $\widehat{s}(u) = (2\pi u - \sin 2\pi u)/2\pi^2 u^2$. Then

$$\begin{aligned} \int_{-\infty}^{+\infty} |h(x)| \, dx &\geq \left| \int_{-\infty}^{+\infty} h(x) s(x) \, dx \right| \\ &= \left| \int_{-\infty}^{+\infty} \widehat{h}(u) \widehat{s}(u) \, du \right| \\ &\geq c \int_1^{+\infty} \frac{\widehat{h}(u)}{u} \, du. \end{aligned}$$

Observe that, through a translation, (3.9) implies the following fact. Suppose $\widehat{h}(u) = 1$ for u in an interval of length r , say $[q, q+r]$, and that, moreover, $\widehat{h}(u) = 0$ for $u \leq q-1$. Then

(3.10)

$$\int_{-\infty}^{+\infty} |h(x)| \, dx \geq c \log r.$$

To prove the lemma we may suppose that B lies in the half-plane $\{(x, y) : x \geq 1\}$ as in Figure 4.

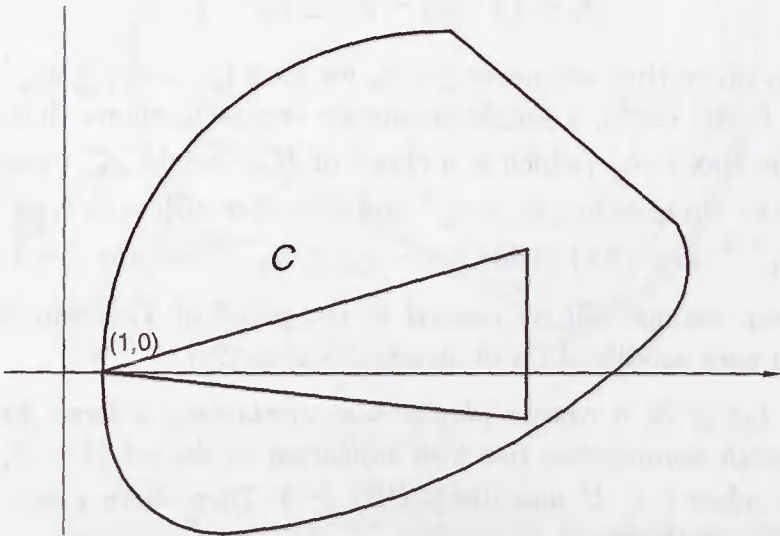


FIGURE 4

Then, by (3.9) and (3.10),

$$\begin{aligned}
 & \int_{\mathbb{R}} \int_{\mathbb{R}} |\widehat{g}(\xi, \eta)| \, d\xi d\eta \\
 &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} g(x, y) e^{-2\pi i \eta y} dy \right\} e^{-2\pi i \xi x} dx \right| d\xi d\eta \\
 &\geq c \int_{\mathbb{R}} \int_1^{+\infty} \frac{1}{x} \left| \int_{\mathbb{R}} g(x, y) e^{-2\pi i \eta y} dy \right| dx d\eta \\
 &\geq c \int_1^r \frac{1}{x} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} g(x, y) e^{-2\pi i \eta y} dy \right| d\eta dx \\
 &\geq c \int_1^r \frac{1}{x} \log x \, dx \\
 &= c \log^2 r,
 \end{aligned}$$

since, because of the convexity of B , we can assume that $g(x, y)$ takes value 1 inside a whole triangle such as the one in the previous picture. \square

Proof of Theorem 10. Arguing by contradiction, we assume the existence of a positive continuous function $\varepsilon(\rho) \rightarrow 0$ (as $\rho \rightarrow +\infty$) such that

$$(3.11) \quad \int_0^{2\pi} |\widehat{\chi}_B(\rho\Theta)| \, d\theta \leq \varepsilon(\rho) \rho^{-2} \log \rho$$

for $\rho \geq 2$. Let φ be a nonnegative radial cutoff function supported in the unit disc; then the convolution

$$g = \chi_{\rho B} * \varphi$$

satisfies the assumptions in the previous lemma (ρB contains a disc of radius $\approx \rho$). Therefore, by (3.11),

$$\begin{aligned}
 \log^2 \rho &\leq c \|\widehat{g}\|_{L^1(\mathbb{R}^2)} \\
 &= c \rho^2 \int_{\mathbb{R}^2} |\widehat{\chi}_B(\rho x) \widehat{\varphi}(x)| \, dx \\
 &\leq c \rho^2 \int_{\mathbb{R}^2} |\widehat{\chi}_B(\rho x)| \frac{1}{1+|x|} \, dx \\
 &\leq c \rho^2 \int_0^{+\infty} \frac{u}{1+u} \int_0^{2\pi} |\widehat{\chi}_B(\rho u \Theta)| \, d\theta du \\
 &= c \int_0^{+\infty} \frac{s}{1+\rho^{-1}s} \int_0^{2\pi} |\widehat{\chi}_B(s\Theta)| \, d\theta ds \\
 &\leq c \left(1 + \int_2^{+\infty} \frac{\varepsilon(s) \log s}{s(1+\rho^{-1}s)} \, ds \right) \\
 &\leq c \left(1 + \int_2^{\rho} \frac{\varepsilon(s) \log s}{s} \, ds + \rho \int_{\rho}^{+\infty} \frac{\varepsilon(s) \log s}{s^2} \, ds \right) \\
 &= A(\rho).
 \end{aligned}$$

To end the proof we observe that

$$\frac{A(\rho)}{\log^2 \rho} \rightarrow 0$$

as $\rho \rightarrow +\infty$, by l'Hôpital's rule. □

Remark 22. Using an induction argument as in [27], the above theorem can be extended to several variables, so that, for any convex body in \mathbb{R}^n ,

$$\limsup_{\rho \rightarrow +\infty} \frac{\rho^n}{\log^{n-1} \rho} \int_{\Sigma_{n-1}} |\widehat{\chi}_B(\rho\sigma)| d\sigma > 0.$$

Remark 23. To prove our theorem we have used an idea introduced in [27] to get lower bounds for Lebesgue constants. Therefore our result shows a relation between the study of Lebesgue constants and the L^1 spherical averages of Fourier transforms of characteristic functions. However, we see no general theorem relating one to the other. See [18] for a related discussion with a number-theoretic flavor.

Remark 24. Estimating $|r(B, \delta, \theta)|$ (see (1.5)) is a geometrical problem which does not necessarily involve the Fourier transform. The previous theorem and the inequality in Lemma 17 imply that, for any convex planar body,

$$\limsup_{\delta \rightarrow 0^+} \frac{1}{\delta \log(1/\delta)} \int_0^{2\pi} |r(B, \delta, \theta)| d\theta > 0.$$

The problem considered in the previous remark could be related to the study of floating bodies (see e.g. [22]), where, in place of fixing δ , one fixes the area

$$(\approx \delta |r(B, \delta, \theta)|)$$

of the small part of B cut away by the chord $r(B, \delta, \theta)$ in the direction Θ .

Proof of Theorem 11. Arguing as in [15] or [7] and applying Theorem 2 and (1.2), we have

$$\begin{aligned} & \int_{\mathbb{T}^2} \int_{SO(2)} |D_\rho(\sigma^{-1}(B) - t)| d\sigma dt \\ &= \rho^2 \int_{\mathbb{T}^2} \int_{SO(2)} \left| \sum_{m \neq 0} \widehat{\chi}_B(\rho\sigma m) e^{2\pi i m \cdot t} \right| d\sigma dt \\ &\leq \rho^2 \int_{\mathbb{T}^2} \int_{SO(2)} \left| \sum_{0 \neq |m| \leq \rho^{(1-2\alpha)/(1+2\alpha)}} \widehat{\chi}_B(\rho\sigma m) e^{2\pi i m \cdot t} \right| d\sigma dt \\ &\quad + \rho^2 \int_{\mathbb{T}^2} \int_{SO(2)} \left| \sum_{|m| > \rho^{(1-2\alpha)/(1+2\alpha)}} \widehat{\chi}_B(\rho\sigma m) e^{2\pi i m \cdot t} \right| d\sigma dt \\ &\leq \rho^2 \sum_{0 \neq |m| \leq \rho^{(1-2\alpha)/(1+2\alpha)}} \int_{SO(2)} |\widehat{\chi}_B(\rho\sigma m)| d\sigma \\ &\quad + \rho^2 \left\{ \int_{SO(2)} \sum_{|m| > \rho^{(1-2\alpha)/(1+2\alpha)}} |\widehat{\chi}_B(\rho\sigma m)|^2 d\sigma \right\}^{1/2} \end{aligned}$$

$$\begin{aligned}
 &\lesssim \rho^2 \sum_{0 \neq |m| \leq \rho^{(1-2\alpha)/(1+2\alpha)}} |\rho m|^{-2+\alpha} \log |\rho m| \\
 &\quad + \rho^2 \left\{ \sum_{|m| > \rho^{(1-2\alpha)/(1+2\alpha)}} |\rho m|^{-3} \right\}^{1/2} \\
 &\lesssim \rho^\alpha \int_1^{\rho^{(1-2\alpha)/(1+2\alpha)}} t^{\alpha-1} \log(\rho t) dt + \rho^{1/2} \left\{ \int_{\rho^{(1-2\alpha)/(1+2\alpha)}}^{+\infty} t^{-2} \right\}^{1/2} \\
 &\lesssim \rho^{2\alpha/(1+2\alpha)}.
 \end{aligned}$$

The lower bound follows from Theorem 2 and the orthogonality argument in [7, p. 269]. \square

Proof of Theorem 16. We prove only the case $0 < \alpha < 1/2$. Write N as a sum of four squares, $N = j^2 + k^2 + \ell^2 + m^2$, and let $a_1, a_2, a_3, a_4 \in [0, 1)$ be pairwise linearly independent on \mathbb{Z} , so that, e.g.,

$$a_1 + \frac{p}{j} \neq a_2 + \frac{q}{k}$$

for any choice of the integers p, q, j, k ($j, k \neq 0$). That is,

$$(3.12) \quad (a_1 + j^{-1}\mathbb{Z}) \cap (a_2 + k^{-1}\mathbb{Z}) = \emptyset$$

when $j \neq k$. Let

$$A_{j^2} = \left\{ \left(a_1 + \frac{p}{j}, \frac{q}{j} \right) \right\}_{p, q \in \mathbb{Z}} \cap \mathbb{T}^2,$$

and let us define $A_{k^2}, A_{\ell^2}, A_{m^2}$ accordingly. Define

$$\mathcal{P} = A_{j^2} \cup A_{k^2} \cup A_{\ell^2} \cup A_{m^2}.$$

By (3.12) \mathcal{P} has cardinality N . Since

$$\begin{aligned}
 &\text{card}(\mathcal{P} \cap B) - N|B| \\
 &= \text{card}(A_{j^2} \cap B) - j^2|B| + \dots + \text{card}(A_{m^2} \cap B) - m^2|B|,
 \end{aligned}$$

it is enough to prove that, say,

$$\int_{\mathbb{T}^2} \int_{SO(2)} |\text{card}(A_{j^2} \cap (\sigma(B) + t)) - j^2|B|| d\theta dt \lesssim N^{\frac{\alpha}{1+2\alpha}} \log N.$$

We can therefore prove the theorem assuming N to be a square, say $N = r^2$, $r \in \mathbb{N}$, and

$$\mathcal{P} = A_N = \left\{ \left(a + \frac{p}{r}, \frac{q}{r} \right) \right\}_{p, q \in \mathbb{Z}^2} \cap U.$$

Now observe that, writing $w = (a, 0)$ and applying Theorem 11, we have

$$\begin{aligned}
& \int_{\mathbb{T}^2} \int_{SO(2)} |D(\mathcal{P}, \theta, t)| dt d\sigma \\
&= \int_{SO(2)} \int_{\mathbb{T}^2} |\text{card}(A_{r^2} \cap (\sigma(B) + t)) - r^2 |B|| dt d\sigma \\
&= \int_{SO(2)} \int_{\mathbb{T}^2} |\text{card}(A_{r^2} \cap (\sigma(B) + t + w)) - r^2 |B|| dt d\sigma \\
&= \int_{SO(2)} \int_{\mathbb{T}^2} \left| \text{card} \left(\left\{ \left(\frac{p}{r}, \frac{q}{r} \right) \right\}_{p,q=0}^{r-1} \cap (\sigma(B) + u) \right) - r^2 |B| \right| du d\sigma \\
&= \int_{SO(2)} \int_{\mathbb{T}^2} |\text{card}(\mathbb{Z}^2 \cap (r\sigma(B) + ru)) - r^2 |B|| du d\sigma \\
&= \int_{SO(2)} \int_{\mathbb{T}^2} |\text{card}(\mathbb{Z}^2 \cap (r\sigma(B) + u)) - r^2 |B|| du d\sigma \\
&\lesssim r^{2\alpha/(1+2\alpha)} \log r \\
&= \frac{1}{2} N^{\alpha/(1+2\alpha)} \log N,
\end{aligned}$$

where we have used the fact that, for a function $f \in L^1(\mathbb{T}^2)$ and for any integer $k \neq 0$,

$$\int_{\mathbb{T}^2} f(ku) du = \int_{\mathbb{T}^2} f(u) du.$$

□

The above argument extends to several variables after replacing the sum of four squares by Hilbert's theorem (Waring's problem).

REFERENCES

- [1] J. Beck, *Irregularities of distribution I*, Acta Math. **159** (1987), 1-49. MR **89c**:11117
- [2] J. Beck and W.W.L. Chen, *Irregularities of distribution*, Cambridge University Press, 1987. MR **88m**:11061
- [3] J. Beck and W.W.L. Chen, *Note on irregularities of distribution II*, Proc. London Math. Soc. **61**(1990), 251-272. MR **91g**:11083
- [4] J. Beck and W.W.L. Chen, *Irregularities of point distribution relative to convex polygons II*, Mathematika **40** (1993), 127-136. MR **94i**:11055
- [5] L. Brandolini and L. Colzani, *Localization and convergence of eigenfunction expansions*, Journal Fourier Anal. Appl. **5** (1999), 431-447. MR **2001g**:42054
- [6] L. Brandolini and L. Colzani, *Decay of Fourier Transforms and Summability of Eigenfunction Expansions*, Ann. Scuola Norm. Sup. Pisa Cl. Sci (4) **29** (2000), 611-638. MR **2002e**:35178
- [7] L. Brandolini, L. Colzani and G. Travaglini, *Average decay of Fourier transforms and integer points in polyhedra*, Ark. Mat. **35** (1997), 253-275. MR **99e**:42021
- [8] L. Brandolini, M. Rigoli and G. Travaglini, *Average decay of Fourier transforms and geometry of convex sets*, Revista Mat. Iberoamericana **14** (1998), 519-560. MR **2000a**:42017
- [9] L. Brandolini and G. Travaglini, *Pointwise convergence of Fejer type means*, Tohoku Math. J. **49** (1997), 323-336. MR **98m**:42010
- [10] J. Bruna, A. Nagel and S. Wainger, *Convex hypersurfaces and Fourier transforms*, Ann. Math. **127** (1988), 333-365. MR **89d**:42023
- [11] W.W.L. Chen, *On irregularities of distribution III*, J. Austr. Math. Soc. **60** (1996), 228-244. MR **97e**:11085
- [12] R. Coifman and G. Weiss, *Extensions of Hardy spaces and their use in analysis*, Bull. Amer. Math. Soc. **83** (1977), 569-645. MR **56**:6264

- [13] L. Colzani, *Fourier transform of distributions in Hardy spaces*, Boll. Un. Mat. Ital. **1** (1982), 403-410. MR **84c**:46042
- [14] H. Davenport, *Note on irregularities of distribution*, Mathematika **1** (1954), 73-79. MR **18**:566a
- [15] D. G. Kendall, *On the number of lattice points inside a random oval*, Quart. J. Math. Oxford Ser. **19** (1948), 1-26. MR **9**:570b
- [16] E. Krätzel, *Lattice points*, Kluwer Academic Publisher, 1988. MR **90e**:11144
- [17] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics, **84**, American Mathematical Society, Providence, RI, 1994. MR **96i**:11002
- [18] F. L. Nazarov and A. N. Podkorytov, *On the behaviour of the Lebesgue constants for two-dimensional Fourier sums over polygons*, St. Petersburg Math. J. **7** (1996), 663-680. MR **96m**:42019
- [19] A. N. Podkorytov, *On the asymptotics of the Fourier transform on a convex curve*, Vestn. Leningrad Univ. Math. **24** (1991), no. 2, 57-65. MR **93h**:42019
- [20] B. Randol, *On the Fourier transform of the indicator function of a planar set*, Trans. Amer. Math. Soc. **139** (1969), 271-278. MR **40**:4678a
- [21] F. Ricci and G. Travaglini, *Convex curves, Radon transforms and convolution operators defined by singular measures*, Proc. Amer. Math. Soc. **129** (2001), 1739-1744. MR **2002i**:42010
- [22] C. Schütt, *The convex floating body and polyhedral approximation*, Israel J. Math. **73** (1991), 65-77. MR **92i**:52009
- [23] A. Seeger and S. Ziesler, *Riesz means associated with convex domains in the plane*, Math. Z. **236** (2001), 643-676. MR **2002j**:42011
- [24] W. T. Sledd and D. A. Stegenga, *An H^1 multiplier theorem*, Ark. Mat. **19** (1981), 265-270. MR **84j**:42018
- [25] M. Tarnopolska-Weiss, *On the number of lattice points in a compact n -dimensional polyhedron*, Proc. Amer. Math. Soc. **74** (1979), 124-127. MR **80c**:10048
- [26] A. A. Yudin and V. A. Yudin, *Polygonal Dirichlet kernels and growth of Lebesgue constants*, Mat. Zametki **37** (1985), 220-236; English transl., Math. Notes **37** (1985), 124-135. MR **86k**:42039
- [27] V. A. Yudin, *Lower bound for Lebesgue constants*, Mat. Zametki **25** (1979), 119-122; English transl., Math. Notes **25** (1979), 63-65. MR **80i**:42010

DIPARTIMENTO DI INGEGNERIA, UNIVERSITÀ DI BERGAMO, VIALE G. MARCONI 5, 24044 DALMINE (BG), ITALY

E-mail address: brandolini@unibg.it

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MISSOURI, COLUMBIA, MISSOURI

E-mail address: iosevich@math.missouri.edu

URL: <http://www.math.missouri.edu/~iosevich/>

DIPARTIMENTO DI MATEMATICA E APPLICAZIONI, UNIVERSITÀ DI MILANO-BICOCCA, VIA BICOCCA DEGLI ARCIMBOLDI 8, 20126 MILANO, ITALY

E-mail address: travaglini@matapp.unimib.it

URL: <http://www.matapp.unimib.it/~travaglini/>

A FREE BOUNDARY PROBLEM FOR A SINGULAR SYSTEM OF DIFFERENTIAL EQUATIONS: AN APPLICATION TO A MODEL OF TUMOR GROWTH

SHANGBIN CUI AND AVNER FRIEDMAN

ABSTRACT. In this paper we consider a free boundary problem for a nonlinear system of two ordinary differential equations, one of which is singular at some points, including the initial point $r = 0$. Because of the singularity at $r = 0$, the initial value problem has a one-parameter family of solutions. We prove that there exists a unique solution to the free boundary problem. The proof of existence employs two “shooting” parameters. Analysis of the profiles of solutions of the initial value problem and tools such as comparison theorems and weak limits of solutions play an important role in the proof. The system considered here is motivated by a model in tumor growth, but the methods developed should be applicable to more general systems.

1. INTRODUCTION

A special feature in tumor growth is proliferation; proliferating cells cause the tumor volume to vary in time, and, as a result, various models developed to describe tumor growth are formulated as initial boundary value problems for partial differential equations with the tumor surface as a free boundary.

It has long been recognized that a tumor contains different populations of cells, such as proliferating cells (i.e., cells that undergo abnormally fast mitosis), necrotic cells (i.e., cells that died because of lack of nutrition) and “in-between” quiescent cells (i.e., cells that are alive but their rate of mitosis is balanced by the rate of natural death).

There are basically two kinds of models in the literature, according to which one of the following two assumptions is adopted:

- (i) different populations are segregated by interface boundaries;
- (ii) different populations are mixed together in different concentrations.

Models with segregated populations were developed by Greenspan [9], [10], Adam [1], Britton and Chaplain [2] and Byrne and Chaplain [3] (see also the references therein). In such models, necrotic cells occupy a central core $r < \rho_1(t)$, proliferating cells occupy an outer layer $\rho_2(t) < r < R(t)$, and quiescent cells reside in the shell $\rho_1(t) < r < \rho_2(t)$; in these regions nutrient and inhibitor concentrations satisfy reaction-diffusion equations. Rigorous analysis of models of this kind was given by

Received by the editors March 15, 2002.

2000 *Mathematics Subject Classification.* Primary 34B15; Secondary 35C10, 35Q80, 92C15.

Key words and phrases. Free boundary problem, stationary solutions, singular differential equations, tumor growth.

Friedman and Reitich [7] and Cui and Friedman [4] (for the case $\rho_1(t) = \rho_2(t) = 0$) and by Cui and Friedman [5] (for the case $\rho_1(t) = \rho_2(t) > 0$).

Models of the type (ii) were more recently developed in Ward and King [16], Pettet *et al.* [13] and Sherratt and Chaplain [14]. In this paper we are particularly interested in the model of Pettet *et al.* [13]. This model was developed in order to explain the experimental results of Dorie *et al.* [6], [7] with regard to internalization of particles injected into the tumor across its surface. Previous models introduced in order to explain the same experimental results were given by McElwain and Pettet [12] and Thompson and Byrne [15].

As in other tumor models, the one developed in [13] considers all nutrients as a single species, and assumes that its concentration C satisfies the diffusion equation

$$(1.1) \qquad \qquad \qquad \nabla^2 C = \mu C,$$

where μ is a positive constant, accounting for the consumption rate of nutrient divided by the diffusion coefficient.

Another basic assumption adopted by [13] is that dead cells are withdrawn immediately from the tumor upon their death, so that the tumor contains only living cells. It is also assumed that all cells are incompressible and of the same volume, and that the tumor is a spheroid well packed with cells. It follows that the cell density within the tumor is a constant, say, N . Thus, denoting by P and Q the densities of the proliferating cells and quiescent cells, respectively (i.e., the numbers of proliferationg cells and quiescent cells per unit volume, respectively), we get

$$(1.2) \qquad \qquad \qquad P + Q = N.$$

Next, proliferating cells are assumed to undergo mitosis at a rate $K_B(C)$ and become quiescent at a rate $K_Q(C)$; the quiescent cells are assumed to revert to proliferating cells at a rate $K_P(C)$ and undergo necrosis at a rate $K_D(C)$. This implies, by the law of conservation of mass, that

$$(1.3) \qquad \frac{\partial P}{\partial t} + \nabla \cdot (\vec{u}_P P) = (K_B(C) - K_Q(C))P + K_P(C)Q,$$

$$(1.4) \qquad \frac{\partial Q}{\partial t} + \nabla \cdot (\vec{u}_Q Q) = K_Q(C)P - (K_D(C) + K_P(C))Q,$$

where \vec{u}_P and \vec{u}_Q are the velocities of proliferating cells and quiescent cells, respectively. The functions $K_B(C)$, $K_D(C)$, $K_P(C)$ and $K_Q(C)$ are taken to be linear functions:

$$(1.5) \qquad \begin{aligned} K_B(C) &= k_B C, & K_D(C) &= k_D (C_0 - C), \\ K_P(C) &= k_P C, & K_Q(C) &= k_Q (C_0 - C), \end{aligned}$$

where C_0 is a positive constant, and the coefficients k_B , k_D , k_P , k_Q are typically given by

$$(1.6) \qquad \qquad \qquad k_P = 0.05, \quad k_D = 0.1, \quad k_B = 1, \quad k_Q = 0.9.$$

The velocities \vec{u}_P and \vec{u}_Q are mutually related by the equation

$$(1.7) \qquad \qquad \qquad \vec{u}_Q = \vec{u}_P + \chi \nabla C,$$

where χ is a chemotactic sensitivity coefficient, assumed to be a nonnegative constant. The last assumption is based on some evidence that proliferating cells seem to be less motile as they undergo mitosis [11].

Introducing the mean velocity

$$\vec{u} = \frac{1}{N}(P\vec{u}_P + Q\vec{u}_Q),$$

we get from (1.3) and (1.4) the equation

$$(1.8) \quad \nabla \cdot \vec{u} = \frac{1}{N}\{K_B(C)P - K_D(C)Q\}.$$

Finally, the movement of the tumor surface is assumed to be governed by the equation of continuity, namely,

$$(1.9) \quad \frac{dR}{dt} = \vec{u} \cdot \vec{n},$$

where $R = R(t)$ is the radius of the tumor spheroid and \vec{n} represents the outward normal to the tumor surface.

Equations (1.1)–(1.9) complemented by appropriate boundary and initial conditions form a model of evolution of a tumor containing only cells that are alive but in two different states (proliferating and quiescent) co-inhabiting the tumor. We are interested in developing a rigorous mathematical treatment of this model. The present paper considers stationary solutions for the case $\chi = 0$; non-stationary solutions and the more difficult case $\chi \neq 0$ will be studied in future work.

After rescaling space and time, and setting

$$p = \frac{P}{N}, \quad q = \frac{Q}{N} = 1 - p, \quad c = \frac{C}{C_0}, \quad \vec{u} = u \frac{x}{|x|},$$

the stationary problem can be reformulated in the following form:

$$(1.10) \quad c'' + \frac{2}{r}c' = \mu c, \quad 0 < r < R,$$

$$(1.11) \quad c'(0) = 0, \quad c(R) = 1,$$

$$(1.12) \quad u' + \frac{2}{r}u = -K_D(c) + K_M(c)p, \quad 0 < r < R,$$

$$(1.13) \quad u(0) = 0,$$

$$(1.14) \quad up' = K_P(c) + (K_M(c) - K_N(c))p - K_M(c)p^2, \quad 0 < r < R,$$

$$(1.15) \quad u(R) = 0,$$

where

$$(1.16) \quad K_M = K_B + K_D, \quad K_N = K_P + K_Q,$$

and $K_B(c)$, $K_D(c)$, $K_P(c)$ and $K_Q(c)$ are rescaled forms of the corresponding functions given by (1.5), that is,

$$(1.17) \quad \begin{aligned} K_B(c) &= C_0 k_{Bc}, & K_D(c) &= C_0 k_D(1 - c), \\ K_P(c) &= C_0 k_{Pc}, & K_Q(c) &= C_0 k_Q(1 - c). \end{aligned}$$

Indeed, equation (1.10) is the radially symmetric form of (1.1). The first boundary condition in (1.11) follows from the radial symmetry of the solution of the differential equation for c , and the second boundary condition means that the tumor receives sufficient nourishment from its host tissue (recall that $K_Q(1) = 0$ and $K_D(1) = 0$, which means that proliferating cells do not become quiescent and quiescent cells do not undergo necrosis when the level of nourishment is $c = 1$). Equation

(1.12) is the radially symmetric version of (1.8) (with q replaced by $1 - p$). If we substitute (1.12) into (1.3) and use the assumption that p is independent of t and $\chi = 0$, we obtain the equation (1.14). The boundary condition (1.13) is a consequence of radial symmetry of \vec{u} . Finally, the free boundary condition (1.15) is the stationary form of (1.9).

The main result of this paper is the following: The stationary problem (1.10)–(1.16) has a unique solution (c, p, u, R) with the properties

$$0 < c(r) < 1, \quad c'(r) > 0, \quad 0 < p(r) < 1, \quad p'(r) > 0, \quad u(r) < 0 \quad \text{for } 0 < r < R.$$

The last inequality means that cells are moving into the interior of the tumor. Actually, our analysis does not depend on the linearity of the functions in (1.17), nor does it depend on the special form of the right-hand side of (1.10). In §2 we shall state our main result in a more general form and then outline the structure of this paper.

We conclude this section by calling attention to the paper of Ward and King [16]. This paper introduces a model for a spheroid tumor with two populations of cells: proliferating cells and dead cells. The model is similar to that of Pettet *et al.* [13] with $\chi = 0$. However, the assumptions made on the coefficients which appear in the equations that are analogous to (1.3), (1.4) are quite different; thus, for example, the function $K_B + K_D$ in [13] is uniformly positive whereas in [16] it changes sign. In [16] the radius $R(t)$ of the tumor is shown, numerically, to increase to infinity at a linear rate, so that one cannot expect to have stationary solutions.

2. THE MAIN RESULT

Consider the following free boundary problem:

$$(2.1) \quad c''(r) + \frac{2}{r}c'(r) = F(c(r)), \quad 0 < r < R,$$

$$(2.2) \quad c'(0) = 0, \quad c(R) = 1,$$

$$(2.3) \quad u'(r) + \frac{2}{r}u(r) = -K_D(c(r)) + K_M(c(r))p(r), \quad 0 < r < R,$$

$$(2.4) \quad u(0) = 0,$$

$$(2.5) \quad u(r)p'(r) = K_P(c(r)) + (K_M(c(r)) - K_N(c(r)))p(r) - K_M(c(r))p^2(r), \quad 0 < r < R,$$

$$(2.6) \quad u(R) = 0,$$

with solution (c, u, p, R) satisfying

$$(2.7) \quad c \in C[0, R] \cap C^2(0, R), \quad u \in C[0, R] \cap C^1(0, R), \quad p \in C[0, R] \cap C^1(0, R)$$

and

$$(2.8) \quad c(r) \geq 0, \quad 0 \leq p(r) \leq 1 \quad \text{for } 0 \leq r \leq R.$$

The functions on the right-hand side of (2.1), (2.3) and (2.5) satisfy the following assumptions:

$$(2.9) \quad F(c), K_D(c), K_M(c), K_N(c) \text{ and } K_P(c) \text{ are analytic in } c, \quad 0 \leq c \leq 1;$$

$$(2.10) \quad F(0) = 0, \quad F'(c) > 0 \quad \text{for } 0 \leq c \leq 1;$$

$$(2.11) \quad \begin{cases} K'_D(c) < 0 \text{ for } 0 \leq c \leq 1, \text{ and } K_D(1) = 0; \\ K'_P(c) > 0 \text{ for } 0 \leq c \leq 1, \text{ and } K_P(0) = 0; \\ K_M(c) = K_B(c) + K_D(c), \text{ where } K_B(c) \text{ satisfies the same} \\ \quad \text{conditions as } K_P(c) \text{ and } K'_B(c) + K'_D(c) > 0 \text{ for } 0 \leq c \leq 1; \\ K_N(c) = K_P(c) + K_Q(c), \text{ where } K_Q(c) \text{ satisfies the same} \\ \quad \text{conditions as } K_D(c). \end{cases}$$

The conditions on $K_B(c)$ and $K_D(c)$ imply that

$$(2.12) \quad K_M(c) > 0, \quad K'_M(c) > 0 \quad \text{for } 0 \leq c \leq 1.$$

Note that the conditions in (2.11) are clearly satisfied for the functions in (1.17); in particular, the inequality $K'_B(c) + K'_D(c) > 0$ follows from the assumption $k_B > k_D$.

The main result of this paper is the following:

Theorem 2.1. *Under the assumptions (2.9)–(2.11), the free boundary problem (2.1)–(2.6) has a unique solution (c, p, u, R) with $R > 0$, c analytic in $[0, R]$, p and u continuous in $[0, R]$ and analytic in $(0, R]$, and*

$$(2.13) \quad 0 < c(r) < 1, \quad c'(r) > 0 \quad \text{if } 0 < r < R, \quad c(0) > 0,$$

$$(2.14) \quad 0 < p(r) < 1, \quad p'(r) > 0 \quad \text{if } 0 < r < R, \quad p(0) > 0,$$

$$(2.15) \quad p(R) = 1, \quad p'(R) > 0,$$

$$(2.16) \quad u(r) < 0 \quad \text{if } 0 < r < R.$$

We shall use the shooting method to prove the existence result. The main idea is as follows:

For each $\lambda \in (0, 1)$ we denote by $c_\lambda(r)$ the solution of (2.1) with initial value

$$(2.17) \quad c_\lambda(0) = \lambda, \quad c'_\lambda(0) = 0.$$

One can easily verify that $c_\lambda(r)$ is analytic for $r \geq 0$, $c'_\lambda(r) > 0$, $\frac{\partial c_\lambda(r)}{\partial \lambda} > 0$ for $r > 0$, $c''_\lambda(0) > 0$, and there exists a unique finite number $R_\lambda > 0$ such that

$$(2.18) \quad c_\lambda(R_\lambda) = 1.$$

Furthermore, $\frac{d}{d\lambda} R_\lambda < 0$ for $0 < \lambda < 1$, $R_\lambda \rightarrow \infty$ if $\lambda \rightarrow 0$, and $R_\lambda \rightarrow 0$ if $\lambda \rightarrow 1$. We substitute $c(r) = c_\lambda(r)$ into the system (2.3)–(2.5) to get

$$(2.19) \quad u'(r) + \frac{2}{r}u(r) = -K_D(c_\lambda(r)) + K_M(c_\lambda(r))p(r), \quad 0 < r < R_\lambda,$$

$$(2.20) \quad u(0) = 0,$$

$$(2.21) \quad u(r)p'(r) = K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda(r)))p(r) - K_M(c_\lambda(r))p^2(r), \\ 0 < r < R_\lambda.$$

Hence the problem of solving the system (2.1)–(2.6) is transformed into the following problem: Find values of λ and corresponding solutions of (2.19)–(2.21) such that the free boundary condition

$$(2.22) \quad u(R_\lambda) = 0$$

is satisfied.

Due to the singularity at $r = 0$ of the differential equations (2.19), (2.21), the shooting method we use in this paper turns out to be quite different from the standard one. Indeed, we shall prove that, unlike other nonsingular ODE problems, solutions of the initial value problem (IVP) (2.19)–(2.21) exhibit the following interesting phenomenon: There exists a critical value $\lambda_\infty \in (0, 1)$ such that, for any $\lambda_\infty < \lambda < 1$, the IVP has a unique solution, while for each $0 < \lambda < \lambda_\infty$, the IVP has a continuum of solutions $(p_{\lambda\psi}, u_{\lambda\psi})$ ($\psi \in \mathbf{R}$). We shall also prove that the value of λ for which the free boundary problem (FBP) (2.19)–(2.22) has a solution must belong to $(0, \lambda_\infty)$. Thus, our “shooting target” has to be reached in two steps: First, for every $\lambda \in (0, \lambda_\infty)$ we need to find a $\bar{\psi} \in \mathbf{R}$ such that the solution $(p_\lambda, u_\lambda) \equiv (p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ of the IVP possesses the “best” approximate properties to the solution of the FBP. Having determined this special solution (p_λ, u_λ) for each $\lambda \in (0, \lambda_\infty)$, we then proceed with the second step of determining a value of λ for which (p_λ, u_λ) satisfies the free boundary condition (2.22).

The structure of the paper is as follows:

In §3 we introduce some auxiliary functions of λ and study their basic properties. These functions play a fundamental role throughout this paper.

In order to prove Theorem 2.1 we first need to solve the IVP near $r = 0$. Since the equations (2.19) and (2.21) (particularly the second one) are singular at $r = 0$, local existence of solutions of the IVP is not ensured by classical results. Sections 4 and 5 are devoted to establishing local existence (i.e., for $0 < r < \delta$, δ small) of solutions for the IVP. In §4 we consider only analytic solutions. We prove that there exists a sequence $\{\lambda_n\}_{n=1}^\infty$ converging to λ_∞ increasingly, such that for each $\lambda \neq \lambda_n$, (2.19)–(2.21) has a unique analytic solution, while for each λ_n , the system has either no analytic solutions or a continuum of analytic solutions.

In §5 we consider the IVP for non-analytic solutions. The key step is to transform the IVP into an equivalent system of integral equations that can be solved by using the contraction mapping principle. The final result is as follows: for each $\lambda_\infty < \lambda < 1$, there is a unique solution, whereas for each $0 < \lambda < \lambda_\infty$, there exists a continuum of classical solutions $(p(r; \lambda, \omega), u(r; \lambda, \omega))$ depending on a real parameter ω .

The family of solutions $(p(r; \lambda, \omega), u(r; \lambda, \omega))$ does not depend continuously on (λ, ω) at the points $\lambda = \lambda_n$. To overcome this difficulty, we introduce, in §6, a parameterization $\omega = \omega(\lambda, \psi)$ with a new parameter $\psi \in \mathbf{R}$ such that the solutions will depend continuously on (λ, ψ) for all $\lambda \in (0, \lambda_\infty)$ and $\psi \in \mathbf{R}$ (for $0 \leq r < \delta$, for some $\delta > 0$). This solution will be denoted by $(p_{\lambda\psi}, u_{\lambda\psi})$.

The next step is to study the profiles of $p_{\lambda\psi}(r)$ and $u_{\lambda\psi}(r)$ and use this information to extend the solution to either the entire interval $[0, R_\lambda]$ or a maximal interval $[0, \bar{R})$ such that $p_{\lambda\psi}(r)$ blows up at $r = \bar{R}$. This is done in §7, where we shall prove, in particular, that $p_{\lambda\psi}(r)$ can change monotonicity (from increasing to decreasing) at most once. It will also be clear from the discussion of this section that the values of λ for which the free boundary condition (2.22) can be satisfied must belong to $(0, \lambda_\infty)$.

In §8 we derive another integral equation formulation of the IVP and use it to introduce the concept of weak solutions. Unlike the integral equation formulation introduced in §5, which holds only locally, the integral equation formulation derived in this section holds globally. It will enable us to work with weak limits of solutions. The main result of this section asserts that a weak solution (p, u) of the IVP is actually a classical solution provided $p \geq 0$.

In §9 we shall perform the first step of the shooting method; that is, we shall prove that for any $\lambda \in (0, \lambda_\infty)$ there is a unique $\bar{\psi}$ such that $(p_\lambda, u_\lambda) \equiv (p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is either a “subsolution” or a “supersolution” of the FBP. By a subsolution of the FBP we mean a solution $(p(r), u(r))$ of the IVP, defined for all $0 \leq r \leq R_\lambda$, that satisfies the conditions

$$0 < p(r) < 1, \quad p'(r) > 0 \quad \text{for } 0 < r < R_\lambda, \\ u(r) < 0 \quad \text{for } 0 < r < r_0, \quad u(r_0) = 0, \quad u(r) > 0 \quad \text{for } r_0 < r \leq R_\lambda,$$

for some $0 < r_0 \leq R_\lambda$. By a supersolution of the FBP we mean a solution $(p(r), u(r))$, defined also for all $0 \leq r \leq R_\lambda$, that satisfies the conditions

$$0 < p(r) < 1, \quad p'(r) > 0, \quad u(r) < 0 \quad \text{for } 0 < r < R_\lambda,$$

and $p(R_\lambda) = 1, u(R_\lambda) < 0$. The discussion of this section shows that subsolutions and supersolutions possess the “best” approximate properties to the solution of the FBP among all solutions of the IVP. We shall also prove that a unique subsolution exists for each λ near λ_∞ .

In §10 we shall perform the second step of the shooting method to get a solution of the FBP. First, we prove that the three sets

$$B_0 \equiv \{\lambda \in (0, \lambda_\infty) : \lambda \text{ corresponds to a solution of the FBP}\}, \\ B_1 \equiv \{\lambda \in (0, \lambda_\infty) : \lambda \text{ corresponds to a subsolution,} \\ \text{but not a solution of the FBP}\}, \\ B_2 \equiv \{\lambda \in (0, \lambda_\infty) : \lambda \text{ corresponds to a supersolution of the FBP}\}$$

do not intersect each other and their union is equal to $(0, \lambda_\infty)$. Next, we prove that B_1 and B_2 are open sets. Finally, we show that every λ near 0 belongs to B_2 , so that $B_2 \neq \emptyset$. Since also $B_1 \neq \emptyset$ (by §9), we conclude that $B_1 \cup B_2 \neq (0, \lambda_\infty)$, so that $B_0 \neq \emptyset$, which means that there exists at least one solution of the free boundary problem. Uniqueness is proved in §11.

Some formulas used in §5 and §6 are proved in the Appendix (§12).

To end this section we want to emphasize that throughout this paper we consider only solutions that satisfy the condition $p(0) \geq 0$; this is of course motivated by the fact that $p(r)$ represents the density of cells. The condition $p(0) \geq 0$ appears implicitly in several places in this paper.

3. AUXILIARY FUNCTIONS

In this section we introduce several functions of λ that will play a fundamental role throughout this paper.

Lemma 3.1. *For any $0 < \lambda \leq 1$ the quadratic equation*

$$(3.1) \quad K_P(\lambda) + (K_M(\lambda) - K_N(\lambda))\alpha - K_M(\lambda)\alpha^2 = 0$$

has a unique positive solution, which we denote by $\alpha(\lambda)$, and

$$0 < \alpha(\lambda) < 1, \quad \alpha'(\lambda) > 0 \quad \text{for } 0 < \lambda < 1, \quad \alpha(1) = 1, \\ \alpha(0+) = \begin{cases} 0 & \text{if } K_D(0) \leq K_Q(0), \\ (K_D(0) - K_Q(0))/K_D(0) & \text{if } K_D(0) > K_Q(0). \end{cases}$$

Proof. Denote by $G(\alpha, \lambda)$ the left-hand side of the equation (3.1). It is clear that for all $0 < \lambda < 1$,

$$\begin{aligned} G(1, \lambda) &= K_P(\lambda) - K_N(\lambda) = -K_Q(\lambda) < 0, \\ G(0, \lambda) &= K_P(\lambda) > 0, \quad \lim_{\alpha \rightarrow \pm\infty} G(\alpha, \lambda) = -\infty. \end{aligned}$$

Hence the equation (3.1) has exactly one positive root in the interval $(0, 1)$. It is given by

(3.2)

$$\alpha(\lambda) = \frac{1}{2K_M(\lambda)} \left(K_M(\lambda) - K_N(\lambda) + \sqrt{(K_M(\lambda) - K_N(\lambda))^2 + 4K_M(\lambda)K_P(\lambda)} \right).$$

By (2.11) and (2.12), it follows that

(3.3)

$$\begin{aligned} \frac{\partial G(\alpha, \lambda)}{\partial \lambda} &= K'_P(\lambda) + (K'_M(\lambda) - K'_N(\lambda))\alpha - K'_M(\lambda)\alpha^2 \\ &= K'_P(\lambda)(1 - \alpha) - K'_Q(\lambda)\alpha + K'_M(\lambda)\alpha(1 - \alpha) > 0 \end{aligned}$$

for $0 < \alpha < 1$, $0 < \lambda < 1$, and, by (3.2),

(3.4)

$$\begin{aligned} \frac{\partial G}{\partial \alpha}(\alpha(\lambda), \lambda) &= (K_M(\lambda) - K_N(\lambda)) - 2K_M(\lambda)\alpha(\lambda) \\ &= -\sqrt{(K_M(\lambda) - K_N(\lambda))^2 + 4K_M(\lambda)K_P(\lambda)} < 0 \end{aligned}$$

for $0 < \lambda < 1$, so that

(3.5)

$$\alpha'(\lambda) = -\frac{\partial G}{\partial \lambda}(\alpha(\lambda), \lambda) / \frac{\partial G}{\partial \alpha}(\alpha(\lambda), \lambda) > 0$$

for $0 < \lambda < 1$. The rest follows immediately from (3.2). □

By the above lemma, it is reasonable to define $\alpha(0) = \alpha(0+)$. From (3.2) one easily finds that $\alpha(\lambda)$ is also differentiable at $\lambda = 0, 1$, and $\alpha'(0) > 0$, $\alpha'(1) > 0$. We introduce the function

(3.6)

$$\beta(\lambda) = \frac{1}{3}(-K_D(\lambda) + K_M(\lambda)\alpha(\lambda)), \quad 0 \leq \lambda \leq 1.$$

Lemma 3.2. *We have $\beta'(\lambda) > 0$ for all $0 \leq \lambda \leq 1$, and there exists a unique number $\lambda_\infty \in (0, 1)$ such that*

(3.7)

$$\beta(\lambda) \begin{cases} < 0 & \text{for } 0 \leq \lambda < \lambda_\infty, \\ = 0 & \text{for } \lambda = \lambda_\infty, \\ > 0 & \text{for } \lambda_\infty < \lambda \leq 1. \end{cases}$$

Proof. Since $K_M(\lambda) = K_B(\lambda) + K_D(\lambda)$ and $K'_B(\lambda) > 0$, $K'_D(\lambda) < 0$, we have

$$\beta'(\lambda) = \frac{1}{3}(-K'_D(\lambda)(1 - \alpha(\lambda)) + K'_B(\lambda)\alpha(\lambda) + K_M(\lambda)\alpha'(\lambda)) > 0$$

for $0 \leq \lambda \leq 1$. Next we note that

$$\beta(0) = -\frac{1}{3}K_D(0)(1 - \alpha(0)) < 0 \quad \text{and} \quad \beta(1) = \frac{1}{3}K_B(1) > 0.$$

Hence there exists a unique $\lambda = \lambda_\infty$ such that (3.7) holds. □

For every integer $n \geq 0$ we introduce the function

(3.8)

$$\begin{aligned} \gamma_n(\lambda) &= n\beta(\lambda) - (K_M(\lambda) - K_N(\lambda)) + 2K_M(\lambda)\alpha(\lambda) \\ &= n\beta(\lambda) + \sqrt{(K_M(\lambda) - K_N(\lambda))^2 + 4K_M(\lambda)K_P(\lambda)}, \end{aligned}$$

where $0 \leq \lambda \leq 1$.

Lemma 3.3. (1) If $\lambda_\infty \leq \lambda \leq 1$, then $\gamma_n(\lambda) > 0$ for all n , and if $0 \leq \lambda < \lambda_\infty$, then

$$(3.9) \quad \gamma_1(\lambda) > \gamma_2(\lambda) > \cdots > \gamma_n(\lambda) > \gamma_{n+1}(\lambda) > \cdots, \\ \lim_{n \rightarrow \infty} \gamma_n(\lambda) = -\infty.$$

(2) There exists a positive integer n_0 such that for every $n \geq n_0$ the following assertions hold:

- (a) $\gamma'_n(\lambda) > 0$ for $0 \leq \lambda \leq 1$,
- (b) the equation $\gamma_n(\lambda) = 0$ has a unique positive root λ_n , and
- (c) the sequence $\{\lambda_n\}_{n=n_0}^\infty$ is monotone increasing and

$$(3.10) \quad \lim_{n \rightarrow \infty} \lambda_n = \lambda_\infty.$$

(3) For $1 \leq n < n_0$, the set of zeros of $\gamma_n(\lambda)$ is either finite or empty.

Proof. Assertion (1) follows immediately from Lemma 3.2. Next we note that $\gamma'_n(\lambda)$ is the sum of $n\beta'(\lambda)$ and a bounded continuous function. Since

$$(3.11) \quad \beta'(\lambda) \geq \text{const.} > 0$$

for $0 \leq \lambda \leq 1$, it follows that (a) holds for sufficiently large n . Thus, for large n , $\gamma_n(\lambda)$ cannot have more than one zero. Since

$$\gamma_n(\lambda_\infty) = \sqrt{(K_M(\lambda_\infty) - K_N(\lambda_\infty))^2 + 4K_M(\lambda_\infty)K_P(\lambda_\infty)} > 0, \quad n = 1, 2, \dots,$$

and, for large n ,

$$\gamma_n(0) = -\frac{n}{3}K_D(0)(1 - \alpha(0)) + |K_D(0) - K_Q(0)| < 0,$$

we see that (b) holds also for sufficiently large n . The assertion that $\{\lambda_n\}_{n=n_0}^\infty$ is monotone increasing follows from (3.9) and the fact that $\gamma'_n(\lambda) > 0$ for large n . To prove (3.10) we note, by (3.8), that

$$\beta(\lambda_n) = -\frac{1}{n}\sqrt{(K_M(\lambda_n) - K_N(\lambda_n))^2 + 4K_M(\lambda_n)K_P(\lambda_n)} \rightarrow 0$$

as $n \rightarrow \infty$. Since $\beta(\lambda_\infty) = 0$ and $\beta'(\lambda) \geq \text{const.} > 0$ for $0 \leq \lambda \leq 1$, the assertion (3.10) follows. Finally, the assertion (3) follows from the fact that $\gamma_n(\lambda)$ is analytic in λ for all $0 \leq \lambda \leq 1$. \square

4. ANALYTIC SOLUTIONS OF THE INITIAL VALUE PROBLEM

In order to solve the free boundary problem (2.19)–(2.22), we first need to investigate general solutions of the initial value problem (2.19)–(2.21). Since, by (2.20), the equation (2.21) is singular at $r = 0$, the existence of solutions of this problem does not follow from a standard theory. To get an insight into the construction of the general solution, we begin by considering solutions that are analytic near $r = 0$.

The main result of this section is the following:

Theorem 4.1. *Let*

$$S = \{\lambda \in (0, 1) : \gamma_n(\lambda) = 0 \text{ for some integer } n \geq 1\}.$$

Then, for any $\lambda \in (0, 1) \setminus S$, there exists a unique analytic solution of (2.19)–(2.21) in some interval $0 \leq r \leq \delta$, $\delta > 0$. If $\lambda \in S$, then the system (2.19)–(2.21) either has no analytic solutions, or it has a 1-parameter family of analytic solutions.

Proof. Since $F(c)$ is analytic in c for $0 \leq c \leq 1$, $c_\lambda(r)$ is analytic in r for $0 \leq r \leq 1$ and $0 < \lambda < 1$. We shall prove that for $\lambda \in (0, 1) \setminus S$ the analytic solution of (2.19)–(2.21) is given by the power series

$$(4.1) \quad u_\lambda(r) = \sum_{n=0}^{\infty} \frac{r^n}{n!} u_\lambda^{(n)}(0), \quad p_\lambda(r) = \sum_{n=0}^{\infty} \frac{r^n}{n!} p_\lambda^{(n)}(0),$$

where the derivatives $u_\lambda^{(n)}(0)$, $p_\lambda^{(n)}(0)$ are computed inductively from (2.19)–(2.21), and they satisfy the inequalities

$$(4.2) \quad |u_\lambda^{(n)}(0)| \leq \frac{H_0 H^{n-1}}{n^2} n!, \quad |p_\lambda^{(n)}(0)| \leq \frac{H_0 H^{n-1}}{n^2} n!$$

for $n \geq 1$ and some positive constants H_0, H (depending on λ); here H_0 is such that (4.2) holds if $n = 1$, and H will be specified later on. Clearly, if these assertions are proved, then the first part of the theorem follows.

We begin by formally computing the derivatives $u_\lambda^{(n)}(0)$, $p_\lambda^{(n)}(0)$. By (2.20) we have $u_\lambda(0) = 0$, so that, by (2.21),

$$K_P(c_\lambda(0)) + (K_M(c_\lambda(0)) - K_N(c_\lambda(0)))p_\lambda(0) - K_M(c_\lambda(0))p_\lambda^2(0) = 0.$$

It follows, by Lemma 3.1, that

$$(4.3) \quad p_\lambda(0) = \alpha(c_\lambda(0)) = \alpha(\lambda),$$

and then, from (2.21),

$$(4.4) \quad u'_\lambda(0) = \frac{1}{3}(-K_D(\lambda) + K_M(\lambda)\alpha(\lambda)) = \beta(\lambda).$$

Next we differentiate the equation (2.21) and take $r = 0$. Since $c'_\lambda(0) = 0$, we get

$$(4.5) \quad \gamma_1(0)p'_\lambda(0) = 0,$$

and since $\gamma_1(\lambda) \neq 0$,

$$(4.6) \quad p'_\lambda(0) = 0.$$

Suppose we have computed $u_\lambda^{(m)}(0)$, $p_\lambda^{(m)}(0)$ for $m = 0, 1, \dots, n-1$. Multiplying (2.19) by r , differentiating n times and taking $r = 0$, we get

$$(4.7) \quad u_\lambda^{(n)}(0) = -\frac{n}{n+2}k_D^{(n-1)}(\lambda) + \frac{n}{n+2} \sum_{m=0}^{n-1} \binom{n-1}{m} k_M^{(m)}(\lambda) p_\lambda^{(n-1-m)}(0),$$

where

$$(4.8) \quad k_i^{(m)}(\lambda) = \left. \frac{d^m}{dr^m} K_i(c_\lambda(r)) \right|_{r=0}, \quad i = D, M, N, P, \quad m = 0, 1, \dots$$

Similarly, by differentiating (2.21) n times and taking $r = 0$, we obtain

$$\begin{aligned}
 \gamma_n(\lambda)p_\lambda^{(n)}(0) &= k_P^{(n)}(\lambda) + \sum_{m=1}^n \binom{n}{m} (k_M^{(m)}(\lambda) - k_N^{(m)}(\lambda)) p_\lambda^{(n-m)}(0) \\
 &\quad - \sum_{m=1}^n \sum_{l=0}^{n-m} \binom{n}{m} \binom{n-m}{l} k_M^{(m)}(\lambda) p_\lambda^{(l)}(0) p_\lambda^{(n-m-l)}(0) \\
 &\quad - K_M(\lambda) \sum_{m=1}^{n-1} \binom{n}{m} p_\lambda^{(m)}(0) p_\lambda^{(n-m)}(0) \\
 &\quad - \sum_{m=2}^n \binom{n}{m} u_\lambda^{(m)}(0) p_\lambda^{(n-m+1)}(0).
 \end{aligned}
 \tag{4.9}$$

Hence $u_\lambda^{(n)}(0)$ and $p_\lambda^{(n)}(0)$ are uniquely determined for all $n \geq 0$ provided $\lambda \notin S$.

Since $\lambda \notin S$, there exists a constant $c_0 > 0$ (depending on λ) such that

$$|\gamma_n(\lambda)| \geq c_0$$

for all $n \geq 1$. Since $c_\lambda(r)$ is analytic in r and $K_i(c)$ is analytic in c , $i = D, M, N, P$, it follows that $K_i(c_\lambda(r))$ is also analytic in r , which implies that

$$|k_i^{(m)}(\lambda)| \leq \frac{A_0 A^m}{m^2} m! \quad (i = D, M, N, P, m = 1, 2, \dots)$$

for some constants A_0, A (depending on λ). Using these inequalities, one can now easily prove that if (4.2) holds for all $1 \leq n \leq m$, then

$$|u_\lambda^{(m+1)}(0)| \leq \frac{CH_0 H^{m-1}}{(m+1)^2} (m+1)!, \quad |p_\lambda^{(m+1)}(0)| \leq \frac{CH_0 H^{m-1}}{(m+1)^2} (m+1)!,$$

where C depends only on A_0, A and H_0 . Taking $H \geq C$, we conclude, by induction, that (4.2) holds for all $n \geq 1$.

Suppose next that $\lambda \in S$. Then there exists a positive integer n such that $\gamma_n(\lambda) = 0$. It follows that (2.19)–(2.21) cannot have an analytic solution if the right-hand side of (4.9) is not equal to zero. If instead the right-hand side of (4.9) vanishes, then we can take an arbitrary value for $p_\lambda^{(n)}(0)$ and argue similarly as above to conclude that (2.19)–(2.21) has an analytic solution for each choice of $p_\lambda^{(n)}(0)$. \square

Remark 4.1. If $\lambda \in S$, then both of the two cases mentioned in Theorem 4.1 can occur. For instance, by (4.5) we see that if $\gamma_1(\lambda) = 0$, then $p'_\lambda(0)$ can be any real number, so that the system (2.19)–(2.21) has a 1-parameter family of analytic solutions. On the other hand, for $n = 2$ the equation (4.9) reads as follows:

$$\begin{aligned}
 \gamma_2(\lambda)p''_\lambda(0) &= \left(K'_P(\lambda) + (K'_M(\lambda) - K'_N(\lambda))\alpha(\lambda) - K'_M(\lambda)\alpha^2(\lambda) \right) c''_\lambda(0) \\
 &= \left(2K_M(\lambda)\alpha(\lambda) - (K_M(\lambda) - K_N(\lambda)) \right) \alpha'(\lambda) c''_\lambda(0) \quad (\text{by (3.3)–(3.5)}) \\
 &= \sqrt{(K_M(\lambda) - K_N(\lambda))^2 + 4K_M(\lambda)K_P(\lambda)} \alpha'(\lambda) c''_\lambda(0) \quad (\text{by (3.2)}).
 \end{aligned}
 \tag{4.10}$$

Since the right-hand side is positive for all $\lambda \in (0, 1)$, the system (2.19)–(2.21) cannot have analytic solutions if $\gamma_2(\lambda) = 0$.

The above examples show that analytic solutions do not depend *continuously* on λ . Therefore, by working with analytic solutions we cannot use the shooting method to find a solution of the free boundary problem (2.19)–(2.22). In the next section we shall consider general (non-analytic) solutions of (2.19)–(2.21) in a small interval $0 \leq r \leq \delta$.

5. NON-ANALYTIC SOLUTIONS OF THE INITIAL VALUE PROBLEM

In this section we shall consider general classical solutions of the problem (2.19)–(2.21). By a *classical solution* of (2.19)–(2.21) on an interval $[0, \delta]$ ($\delta > 0$) we mean a pair of functions p, u in $C[0, \delta] \cap C^1(0, \delta)$ that satisfy the equations (2.19), (2.21) for $0 < r < \delta$, with $u(0) = 0$.

We first consider solutions of (2.19)–(2.21) in the class

$$(5.1) \quad u \in C^1[0, \delta], \quad p \in C^1[0, \delta]$$

for some $\delta > 0$. By (2.21) and (2.19), for such solutions we have

$$(5.2) \quad p(0) = \alpha(\lambda), \quad u'(0) = \beta(\lambda).$$

We shall also impose the condition

$$(5.3) \quad p'(0) = 0.$$

This condition is satisfied if $p \in C^2[0, \delta]$ and $\gamma_1(\lambda) \neq 0$ (but see Theorem 5.4).

We introduce new variables P, U by

$$(5.4) \quad p(r) = \alpha(\lambda) + rP(r), \quad u(r) = r(\beta(\lambda) + U(r)).$$

Then

$$(5.5) \quad P \in C[0, \delta] \cap C^1(0, \delta], \quad U \in C[0, \delta] \cap C^1(0, \delta],$$

$$(5.6) \quad P(0) = 0, \quad U(0) = 0,$$

and

$$(5.7) \quad U'(r) + \frac{3}{r}U(r) = K_M(c_\lambda(r))P(r) + m_\lambda(r),$$

$$(5.8) \quad P'(r) - \frac{\sigma(\lambda)}{r}P(r) = f_\lambda(r, P(r), U(r)),$$

where

$$(5.9) \quad \sigma(\lambda) = -\frac{\gamma_1(\lambda)}{\beta(\lambda)},$$

$$(5.10) \quad m_\lambda(r) = \frac{1}{r}(-K_D(c_\lambda(r)) + K_M(c_\lambda(r))\alpha(\lambda) - 3\beta(\lambda)),$$

$$(5.11) \quad f_\lambda(r, P, U) = \left(-K_M(c_\lambda(r))P^2 + \frac{g_\lambda(0)}{\beta(\lambda)} \frac{UP}{r} - \frac{g_\lambda(r) - g_\lambda(0)}{r} P + \frac{h_\lambda(r)}{r^2} \right) / (U + \beta(\lambda)),$$

where

$$(5.12) \quad g_\lambda(r) = 2K_M(c_\lambda(r))\alpha(\lambda) - (K_M(c_\lambda(r)) - K_N(c_\lambda(r))),$$

$$(5.13) \quad h_\lambda(r) = K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda(r)))\alpha(\lambda) - K_M(c_\lambda(r))\alpha^2(\lambda).$$

Since the function $-K_D(c_\lambda(r)) + K_M(c_\lambda(r))\alpha(\lambda) - 3\beta(\lambda)$ is in C^∞ up to $r = 0$ and is equal to zero at $r = 0$, the function $m_\lambda(r)$ is in C^∞ up to $r = 0$. Similarly, the functions $(g_\lambda(r) - g_\lambda(0))/r$ and $h_\lambda(r)/r^2$ are also in C^∞ up to $r = 0$. Furthermore,

$$(5.14) \quad m_\lambda(r) = \frac{r}{2} c_\lambda''(0) (-K'_D(\lambda) + K'_M(\lambda)\alpha(\lambda)) + O(r^2),$$

$$(5.15) \quad \frac{1}{r}(g_\lambda(r) - g_\lambda(0)) = \frac{r}{2} c_\lambda''(0) \{2K'_M(\lambda)\alpha(\lambda) - (K'_M(\lambda) - K'_N(\lambda))\} + O(r^2),$$

$$(5.16) \quad \frac{h_\lambda(r)}{r^2} = \frac{1}{2} c_\lambda''(0) \{K'_P(\lambda) + (K'_M(\lambda) - K'_N(\lambda))\alpha(\lambda) - K'_M(\lambda)\alpha^2(\lambda)\} + O(r).$$

We shall later on use the inequalities

$$\sigma(\lambda) > -1 \quad \text{if } \lambda \in (0, \lambda_\infty); \quad \sigma(\lambda) < -1 \quad \text{if } \lambda \in (\lambda_\infty, 1).$$

We want to recast the differential equations for U , P as integral equations. Clearly, equation (5.7) with the condition $U(0) = 0$ is equivalent to the integral equation

$$(5.17) \quad U(r) = \frac{1}{r^3} \int_0^r K_M(c_\lambda(\rho)) P(\rho) \rho^3 d\rho + \frac{1}{r^3} \int_0^r m_\lambda(\rho) \rho^3 d\rho.$$

In order to derive an integral equation for P , we first introduce some new notation.

For any $\lambda \in [0, 1] \setminus S$ we define inductively $\alpha_n(\lambda)$ ($n \geq 0$) and $\beta_n(\lambda)$ ($n \geq 1$) as follows:

$$(5.18) \quad \alpha_0(\lambda) = \alpha(\lambda), \quad \beta_1(\lambda) = \beta(\lambda),$$

(5.19)

$$\begin{aligned} \alpha_n(\lambda) = & \frac{1}{\gamma_n(\lambda)} \left\{ k_P^{(n)}(\lambda) + \sum_{m=1}^n \binom{n}{m} (k_M^{(m)}(\lambda) - k_N^{(m)}(\lambda)) \alpha_{n-m}(\lambda) \right. \\ & - \sum_{m=1}^n \sum_{l=0}^{n-m} \binom{n}{m} \binom{n-m}{l} k_M^{(m)}(\lambda) \alpha_l(\lambda) \alpha_{n-m-l}(\lambda) \\ & \left. - K_M(\lambda) \sum_{m=1}^{n-1} \binom{n}{m} \alpha_m(\lambda) \alpha_{n-m}(\lambda) - \sum_{m=2}^n \binom{n}{m} \beta_m(\lambda) \alpha_{n-m+1}(\lambda) \right\}, \end{aligned}$$

$$(5.20) \quad \beta_{n+1}(\lambda) = -\frac{n+1}{n+3} k_D^{(n)}(\lambda) + \frac{n+1}{n+3} \sum_{m=0}^n \binom{n}{m} k_M^{(m)}(\lambda) \alpha_{n-m}(\lambda),$$

for $n = 1, 2, \dots$, where $k_i^{(n)}(\lambda)$ are as in (4.8). Note that

$$\alpha_n(\lambda) = p_\lambda^{(n)}(0), \quad \beta_n(\lambda) = u_\lambda^{(n)}(0) \quad (\beta_0(\lambda) = 0),$$

for $0 < \lambda < 1$, where (p_λ, u_λ) is the unique analytic solution of (2.19)–(2.21) (cf. (4.7), (4.9)). Note also, by (5.3) and the fact that $c'_\lambda(0) = 0$, that

$$\alpha_1(\lambda) = 0, \quad \beta_2(\lambda) = 0.$$

We also define

$$(5.21) \quad \mu_n(\lambda) \equiv \frac{\alpha_{n+2}(\lambda) \gamma_{n+2}(\lambda)}{(n+2)! \beta(\lambda)} = \frac{\alpha_{n+2}(\lambda)}{(n+2)!} (n+1 - \sigma(\lambda)) \quad (n = 0, 1, 2, \dots).$$

For $\lambda \in S$, $\gamma_1(\lambda) \neq 0$, if n is the nonnegative integer such that $\gamma_{n+2}(\lambda) = 0$ (i.e., $\sigma(\lambda) = n + 1$), then $\alpha_i(\lambda)$ ($0 \leq i \leq n + 1$), $\beta_i(\lambda)$ ($1 \leq i \leq n + 2$) and $\mu_i(\lambda)$ ($0 \leq i \leq n - 1$) are still well-defined by (5.19)–(5.21). We also define $\mu_n(\lambda)$ by

$$(5.22) \quad \begin{aligned} \mu_n(\lambda) = & \frac{1}{(n+2)!\beta(\lambda)} \left\{ k_P^{(n+2)}(\lambda) + \sum_{m=1}^{n+2} \binom{n+2}{m} (k_M^{(m)}(\lambda) - k_N^{(m)}(\lambda)) \alpha_{n-m+2}(\lambda) \right. \\ & - \sum_{m=1}^{n+2} \sum_{l=0}^{n-m+2} \binom{n+2}{m} \binom{n+2-m}{l} k_M^{(m)}(\lambda) \alpha_l(\lambda) \alpha_{n-m-l+2}(\lambda) \\ & - K_M(\lambda) \sum_{m=1}^{n+1} \binom{n+2}{m} \alpha_m(\lambda) \alpha_{n-m+2}(\lambda) \\ & \left. - \sum_{m=2}^{n+2} \binom{n+2}{m} \beta_m(\lambda) \alpha_{n-m+3}(\lambda) \right\}. \end{aligned}$$

This is consistent with (5.21), since $\gamma_{n+2}(\lambda)$ appears as a denominator in the definition of $\alpha_{n+2}(\lambda)$ (see (5.19)).

The following lemma gives an equivalent integral form of the equation (5.8).

Lemma 5.1. *Let $P(r)$, $U(r) \in C[0, \delta] \cap C^1(0, \delta]$ be a solution of (5.6)–(5.8) and assume that $\lambda \neq \lambda_\infty$. Then, if δ is sufficiently small, the following assertions hold:*

(1) *If either $\beta(\lambda) > 0$ or $\gamma_1(\lambda) \leq 0$ ($\iff \sigma(\lambda) \leq 0$), then $(P(r), U(r))$ satisfies the equation*

$$(5.23) \quad P(r) = r^{\sigma(\lambda)} \int_0^r f_\lambda(\rho, P(\rho), U(\rho)) \rho^{-\sigma(\lambda)} d\rho.$$

(2) *If $\gamma_1(\lambda) > 0$ and $\gamma_2(\lambda) < 0$ ($\iff 0 < \sigma(\lambda) < 1$), then the limit*

$$\omega = \lim_{r \rightarrow 0} r^{-\sigma(\lambda)} P(r)$$

exists, and $(P(r), U(r))$ satisfies the equation

$$(5.24) \quad P(r) = \omega r^{\sigma(\lambda)} + r^{\sigma(\lambda)} \int_0^r f_\lambda(\rho, P(\rho), U(\rho)) \rho^{-\sigma(\lambda)} d\rho.$$

(3) *If $\gamma_2(\lambda) = 0$ ($\iff \sigma(\lambda) = 1$), then the limit*

$$\omega = \lim_{r \rightarrow 0} r^{-1} (P(r) - \mu_0(\lambda) r \log r)$$

exists, and $(P(r), U(r))$ satisfies the equation

$$(5.25) \quad P(r) = \omega r + \mu_0(\lambda) r \log r + r \int_0^r (f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda)) \rho^{-1} d\rho.$$

(4) *In general, if for some integer $n \geq 2$ we have $\gamma_n(\lambda) > 0$, $\gamma_{n+1}(\lambda) < 0$ ($\iff n - 1 < \sigma(\lambda) < n$), then the limit*

$$\omega = \lim_{r \rightarrow 0} r^{-\sigma(\lambda)} \left(P(r) - \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^m \right)$$

exists, and $(P(r), U(r))$ satisfies the equation

$$(5.26) \quad \begin{aligned} P(r) = & \omega r^{\sigma(\lambda)} + \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^m \\ & + r^{\sigma(\lambda)} \int_0^r \left(f_\lambda(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-2} \mu_m(\lambda) \rho^m \right) \rho^{-\sigma(\lambda)} d\rho. \end{aligned}$$

If instead $\gamma_{n+1}(\lambda) = 0$ ($\iff \sigma(\lambda) = n$), then the limit

$$\omega = \lim_{r \rightarrow 0} r^{-n} \left(P(r) - \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^m - \mu_{n-1}(\lambda) r^n \log r \right)$$

exists, and $(P(r), U(r))$ satisfies the equation

$$(5.27) \quad \begin{aligned} P(r) = & \omega r^n + \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^m + \mu_{n-1}(\lambda) r^n \log r \\ & + r^n \int_0^r \left(f_\lambda(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-2} \mu_m(\lambda) \rho^m \right) \rho^{-n} d\rho. \end{aligned}$$

Conversely, if $P(r), U(r) \in C[0, \delta]$ and they satisfy the coupled system of equations (5.17) and one of the equations (5.23)–(5.27) (in accordance with the corresponding condition on λ), then $(P(r), U(r))$ is a solution of the problem (5.5)–(5.8).

We need the following lemma:

Lemma 5.2. (1) Let $P(r), U(r) \in C[0, \delta]$ and suppose that they satisfy the equation (5.17). Suppose further that, for some integer $n \geq 1$,

$$(5.28) \quad P(r) = \sum_{m=1}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} + O(r^n) \quad \text{as } r \rightarrow 0.$$

Then

$$(5.29) \quad U(r) = \sum_{m=2}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} + O(r^{n+1}) \quad \text{as } r \rightarrow 0.$$

(2) Let $f_\lambda(r, P(r), U(r))$ be the function given by (5.11) and $\lambda \neq \lambda_\infty$. If $P(r), U(r)$ have the expansions (5.28) and (5.29), then

$$(5.30) \quad f_\lambda(r, P(r), U(r)) = \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + O(r^n) \quad \text{as } r \rightarrow 0.$$

Proof. Since the functions $K_M(c_\lambda(r))$ and $m_\lambda(r)$ are in C^∞ up to $r = 0$, we have

$$K_M(c_\lambda(r)) = \sum_{m=0}^n \frac{1}{m!} k_M^{(m)}(\lambda) r^m + O(r^{n+1}),$$

as $r \rightarrow 0$, and a similar expansion holds for $m_\lambda(r)$. Substituting these expressions and (5.28) into (5.17) yields (5.29). Next, since $(g_\lambda(r) - g_\lambda(0))/r$ and $h_\lambda(r)/r^2$ are also in C^∞ up to $r = 0$ and $U(0) + \beta(\lambda) \neq 0$, it is clear that $f_\lambda(r, P(r), U(r))$ has

an expansion

$$f_\lambda(r, P(r), U(r)) = \sum_{m=0}^{n-1} a_m(\lambda) r^m + O(r^n) \text{ as } r \rightarrow 0.$$

The proof that $a_m(\lambda) = \mu_m(\lambda)$ is given in the Appendix (§12). \square

Proof of Lemma 5.1. Consider Case (1) first. It is clear that $f_\lambda(r, P(r), U(r))$ is bounded for $r \in (0, \delta]$ if δ is sufficiently small. Thus multiplying (5.8) by $r^{-\sigma(\lambda)}$ and integrating over the arbitrary interval $[0, r]$ immediately yields the equation (5.23).

Consider Case (2) next. Since $P(r) \in C[0, \delta]$, there exists a constant C (depending on λ and δ) such that

$$(5.31) \quad |P(r)| \leq C \quad \text{for } 0 \leq r \leq \delta.$$

In the sequel we shall use C to denote various positive constants (depending on λ and δ). By (5.17) it follows that

$$(5.32) \quad |U(r)| \leq Cr \quad \text{for } 0 \leq r \leq \delta.$$

Thus $U(r)/r$ is bounded, and

$$|U(r) + \beta(\lambda)| \geq C > 0 \quad \text{for } 0 \leq r \leq \delta$$

if δ is sufficiently small. Consequently,

$$(5.33) \quad |f_\lambda(r, P(r), U(r))| \leq C \quad \text{for } 0 < r \leq \delta$$

and, since $0 < \sigma(\lambda) < 1$, $f_\lambda(r, P(r), U(r))r^{-\sigma(\lambda)}$ is integrable on $(0, \delta]$. We now multiply (5.8) by $r^{-\sigma(\lambda)}$ and integrate over the interval $[s, r]$, for arbitrary $0 < s < r \leq \delta$, to get

$$(5.34) \quad r^{-\sigma(\lambda)}P(r) - s^{-\sigma(\lambda)}P(s) = \int_s^r f_\lambda(\rho, P(\rho), U(\rho))\rho^{-\sigma(\lambda)}d\rho.$$

Since the right-hand side converges to zero as $r \rightarrow 0$, the limit

$$\omega = \lim_{r \rightarrow 0} r^{-\sigma(\lambda)}P(r)$$

exists. Multiplying (5.34) by $r^{-\sigma(\lambda)}$ and letting $s \rightarrow 0$, (5.24) follows.

Consider Case (3). We rewrite (5.34) in the form

$$(5.35) \quad r^{-1}P(r) - s^{-1}P(s) = \int_s^r (f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda))\rho^{-1}d\rho + \mu_0(\lambda)(\log r - \log s).$$

We claim that

$$(5.36) \quad |P(r)| \leq Cr \log \frac{1}{r} \quad \text{for } 0 < r \leq \delta.$$

Indeed, as before, (5.31)–(5.34) hold (with $\sigma(\lambda) = 1$), so that

$$|\delta^{-1}P(\delta) - r^{-1}P(r)| \leq \int_r^\delta |f_\lambda(\rho, P(\rho), U(\rho))|\rho^{-1}d\rho \leq C(\log \delta - \log r)$$

for $0 < r \leq \delta$, and (5.36) follows. Substituting (5.36) into (5.17), we find that

$$(5.37) \quad |U(r)| \leq Cr^2 \log \frac{1}{r} \quad \text{for } 0 < r \leq \delta.$$

It follows that

$$\begin{aligned}
 & |f_\lambda(r, P(r), U(r)) - \mu_0(\lambda)| \\
 (5.38) \quad & \leq C|P(r)|^2 + C\left|\frac{U(r)P(r)}{r}\right| + Cr|P(r)| + \left|\frac{h_\lambda(r)}{r^2(U(r) + \beta(\lambda))} - \mu_0(\lambda)\right| \\
 & \leq Cr^2 \log^2 r + Cr^2 \log^2 r + Cr^2 \log \frac{1}{r} + Cr \leq Cr,
 \end{aligned}$$

so that $\int_0^\delta |f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda)| \rho^{-1} d\rho < \infty$. Consequently, by (5.35), the limit

$$\omega = \lim_{r \rightarrow 0} r^{-1}(P(r) - \mu_0(\lambda)r \log r)$$

exists, and letting $s \rightarrow 0$ in (5.35) we see that (5.25) holds.

Consider now the general case $\gamma_n(\lambda) > 0$, $\gamma_{n+1}(\lambda) < 0$, namely, $n-1 < \sigma(\lambda) < n$, where $n \geq 2$. As before, (5.31)–(5.34) hold, so that

$$\begin{aligned}
 |\delta^{-\sigma(\lambda)} P(\delta) - r^{-\sigma(\lambda)} P(r)| & \leq \int_r^\delta |f_\lambda(\rho, P(\rho), U(\rho))| \rho^{-\sigma(\lambda)} d\rho \\
 & \leq C(\delta^{1-\sigma(\lambda)} - r^{1-\sigma(\lambda)})
 \end{aligned}$$

for $0 < r \leq \delta$. It follows that

$$(5.39) \quad |P(r)| \leq Cr \quad \text{for } 0 \leq r \leq \delta.$$

By (5.17) we further deduce that

$$(5.40) \quad |U(r)| \leq Cr^2 \quad \text{for } 0 \leq r \leq \delta.$$

As in the derivation of (5.38), these estimates allow us to prove that

$$(5.41) \quad |f_\lambda(r, P(r), U(r)) - \mu_0(\lambda)| \leq Cr \quad \text{for } 0 < r \leq \delta.$$

We rewrite (5.34) in the form

$$\begin{aligned}
 (5.42) \quad & r^{-\sigma(\lambda)} P(r) - s^{-\sigma(\lambda)} P(s) \\
 & = \int_s^r (f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda)) \rho^{-\sigma(\lambda)} d\rho + \frac{\mu_0(\lambda)}{1-\sigma(\lambda)} (r^{1-\sigma(\lambda)} - s^{1-\sigma(\lambda)}).
 \end{aligned}$$

If $1 < \sigma(\lambda) < 2$, then $\int_0^\delta |f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda)| \rho^{-\sigma(\lambda)} d\rho < \infty$, and the desired assertion follows by an argument similar to that used in the case $\sigma(\lambda) = 1$. It remains to consider the case $\sigma(\lambda) > 2$. By (5.42) and (5.41) it follows that

$$\begin{aligned}
 & \left| \left(\delta^{-\sigma(\lambda)} P(\delta) - r^{-\sigma(\lambda)} P(r) \right) - \frac{\mu_0(\lambda)}{1-\sigma(\lambda)} (\delta^{1-\sigma(\lambda)} - r^{1-\sigma(\lambda)}) \right| \\
 & \leq \int_r^\delta |f_\lambda(\rho, P(\rho), U(\rho)) - \mu_0(\lambda)| \rho^{-\sigma(\lambda)} d\rho \\
 & \leq C|\delta^{2-\sigma(\lambda)} - r^{2-\sigma(\lambda)}|
 \end{aligned}$$

for $0 < r \leq \delta$. Hence we have

$$\left| P(r) - \frac{\mu_0(\lambda)}{1-\sigma(\lambda)} r \right| \leq Cr^2 \quad \text{for } 0 \leq r \leq \delta,$$

or equivalently (by (5.21)),

$$\left| P(r) - \frac{1}{2} \alpha_2(\lambda) r \right| \leq Cr^2 \quad \text{for } 0 \leq r \leq \delta.$$

It follows, by Lemma 5.2 (1), that

$$\left| U(r) - \frac{1}{6}\beta_3(\lambda)r^2 \right| \leq Cr^3 \quad \text{for } 0 \leq r \leq \delta,$$

and by Lemma 5.2 (2), that

$$|f_\lambda(r, P(r), U(r)) - \mu_0(\lambda) - \mu_1(\lambda)r| \leq Cr^2 \quad \text{for } 0 < r \leq \delta.$$

Repeating the above bootstrap argument step-by-step, we arrive at the estimate

$$\left| f_\lambda(r, P(r), U(r)) - \sum_{m=0}^{n-2} \mu_m(\lambda)r^m \right| \leq Cr^{n-1} \quad \text{for } 0 < r \leq \delta,$$

so that

$$\int_0^\delta \left| f_\lambda(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-2} \mu_m(\lambda)\rho^m \right| \rho^{-\sigma(\lambda)} d\rho < \infty.$$

Since for any $0 < s < r \leq \delta$ we have

$$\begin{aligned} & \left(r^{-\sigma(\lambda)} P(r) - \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^{m-\sigma(\lambda)} \right) - \left(s^{-\sigma(\lambda)} P(s) - \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} s^{m-\sigma(\lambda)} \right) \\ &= \int_s^r \left(f_\lambda(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-2} \mu_m(\lambda)\rho^m \right) \rho^{-\sigma(\lambda)} d\rho, \end{aligned}$$

it follows that the limit

$$\omega = \lim_{r \rightarrow 0} r^{-\sigma(\lambda)} \left(P(r) - \sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{(m+1)!} r^m \right)$$

exists and, by a similar argument as before, (5.26) holds.

The argument for the case $\gamma_{n+1}(\lambda) = 0$ (i.e., $\sigma(\lambda) = n$) is similar.

Finally, it is rather immediate to verify that if $(P(r), U(r))$ satisfies (5.17) and one of the equations (5.23)–(5.27), then it also satisfies (5.6)–(5.8). \square

The first main result of this section is as follows:

Theorem 5.3. *Let $\lambda \neq \lambda_\infty$. Then the following hold:*

(1) *If either $\beta(\lambda) > 0$ or $\gamma_1(\lambda) \leq 0$ ($\iff \sigma(\lambda) \leq 0$), then the system (2.19)–(2.21) has a unique local solution satisfying conditions (5.1)–(5.3).*

(2) *If $\beta(\lambda) < 0$ and $\gamma_1(\lambda) > 0$ ($\iff \sigma(\lambda) > 0$), then the system (2.19)–(2.21) has a 1-parameter family of local solutions satisfying conditions (5.1)–(5.3). More precisely, we have:*

(i) *If $\gamma_n(\lambda) > 0$, $\gamma_{n+1}(\lambda) < 0$ ($\iff n-1 < \sigma(\lambda) < n$) for some positive integer n , then for every real number ω the above system has a unique local solution satisfying the following conditions:*

$$(5.43) \quad p(r) = \sum_{m=0}^n \frac{\alpha_m(\lambda)}{m!} r^m + \omega r^{1+\sigma(\lambda)} + O(r^{n+1}) \quad \text{as } r \rightarrow 0,$$

$$(5.44) \quad u(r) = \sum_{m=1}^{n+1} \frac{\beta_m(\lambda)}{m!} r^m + \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} r^{2+\sigma(\lambda)} + O(r^{n+2}) \quad \text{as } r \rightarrow 0;$$

moreover, these are all the solutions of (2.19)–(2.21) satisfying (5.1)–(5.3).

(ii) If $\gamma_{n+1}(\lambda) = 0$ ($\iff \sigma(\lambda) = n$) for some positive integer n , then for every real number ω the above system has a unique local solution satisfying the following conditions:

(5.45)

$$p(r) = \sum_{m=0}^n \frac{\alpha_m(\lambda)}{m!} r^m + \mu_{n-1}(\lambda) r^{n+1} \log r + \omega r^{n+1} + O(r^{n+2} \log r) \text{ as } r \rightarrow 0,$$

$$(5.46) \quad u(r) = \sum_{m=1}^{n+1} \frac{\beta_m(\lambda)}{m!} r^m + \frac{\mu_{n-1}(\lambda) K_M(\lambda)}{n+4} r^{n+2} \log r + \tilde{\omega} r^{n+2} + O(r^{n+3} \log r) \\ \text{as } r \rightarrow 0,$$

where

$$\tilde{\omega} = \frac{K_M(\lambda)}{n+4} \left(\omega - \frac{\mu_{n-1}(\lambda)}{n+4} \right) \\ + \frac{1}{(n+4)(n+1)!} \left(-k_D^{(n+1)}(\lambda) + \sum_{m=0}^n \binom{n+1}{m} k_M^{(n+1)}(\lambda) \alpha_m(\lambda) \right);$$

moreover, these are all the solutions of (2.19)–(2.21) satisfying (5.1)–(5.3).

Proof. By Lemma 5.1, we only need to prove that the system of equations formed by (5.17) coupled with one of the equations (5.23)–(5.27) has a unique solution in the class $C[0, \delta]$, for some small $\delta > 0$. We shall use the contraction mapping principle to prove this.

The argument for the system of equations (5.17), (5.23) is simple, and we omit it. Since (5.24) and (5.25) are special situations of (5.26) and (5.27), respectively, it suffices to consider the systems (5.17), (5.26) and (5.17), (5.27).

We first consider the system (5.17), (5.26). For small $\delta > 0$, we denote by \tilde{X}_δ the set of continuous functions $(P(r), U(r))$ defined on $[0, \delta]$ satisfying the following conditions:

$$(5.47) \quad \left| P(r) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} \right| \leq (1 + |\omega|) r^{\sigma(\lambda)} \text{ for } 0 \leq r \leq \delta,$$

$$(5.48) \quad \left| U(r) - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} \right| \leq \left(1 + \left| \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} \right| \right) r^{1+\sigma(\lambda)} \text{ for } 0 \leq r \leq \delta.$$

As in the proof of Lemma 5.2, one can show that if δ is small enough, then

$$(5.49) \quad \left| f_\lambda(r, P(r), U(r)) - \sum_{m=0}^{n-2} \mu_m(\lambda) r^m \right| \leq C r^{n-1}$$

for $0 < r \leq \delta$ and for all $(P(r), U(r)) \in \tilde{X}_\delta$. Here and later on we use C to denote various positive constants depending only on δ , λ , ω and n (but independent of the specific functions (P, U)). We now introduce a metric space (X_δ, d) as follows: The set X_δ consists of continuous functions $(P(r), U(r))$ defined on $[0, \delta]$, satisfying the inequalities

$$(5.50) \quad \left| P(r) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} - \omega r^{\sigma(\lambda)} \right| \leq M_1 r^n,$$

$$(5.51) \quad \left| U(r) - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} - \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} r^{1+\sigma(\lambda)} \right| \leq M_2 r^{n+1},$$

for $0 \leq r \leq \delta$, where $M_1 = C/(n - \sigma(\lambda))$, C being the constant in (5.49), and M_2 is another constant to be specified later on. The distance function d is defined by

$$d((P_1, U_1), (P_2, U_2)) = \sup_{0 < r \leq \delta} \frac{|P_1(r) - P_2(r)|}{r^{\sigma(\lambda)}} + \sup_{0 < r \leq \delta} \frac{|U_1(r) - U_2(r)|}{r^{1+\sigma(\lambda)-\theta}},$$

where $0 < \theta < 1$ is an arbitrarily chosen number. Clearly, (X_δ, d) is a complete metric space, and $X_\delta \subset \tilde{X}_\delta$ if δ is sufficiently small.

Consider the mapping $\mathcal{F} : (P, U) \rightarrow (\tilde{P}, \tilde{U})$ defined by

$$(5.52) \quad \begin{aligned} \tilde{P}(r) = & \omega r^{\sigma(\lambda)} + \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} \\ & + r^{\sigma(\lambda)} \int_0^r \left(f_\lambda(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-2} \mu_m(\lambda) \rho^m \right) \rho^{-\sigma(\lambda)} d\rho, \end{aligned}$$

$$(5.53) \quad \tilde{U}(r) = \frac{1}{r^3} \int_0^r K_M(c_\lambda(\rho)) P(\rho) \rho^3 d\rho + \frac{1}{r^3} \int_0^r m_\lambda(\rho) \rho^3 d\rho,$$

where $(P, U) \in \tilde{X}_\delta$. We claim that $(\tilde{P}, \tilde{U}) \in X_\delta$. Indeed, using (5.49), it follows immediately that $\tilde{P}(r)$ satisfies (5.50). To see that $\tilde{U}(r)$ satisfies (5.51) we set

$$\begin{aligned} \Delta_\lambda(r) \equiv & \frac{1}{r^3} \int_0^r K_M(c_\lambda(\rho)) \left(\sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} \rho^{m-1} + \omega \rho^{\sigma(\lambda)} \right) \rho^3 d\rho \\ & + \frac{1}{r^3} \int_0^r m_\lambda(\rho) \rho^3 d\rho - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} - \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} r^{1+\sigma(\lambda)}. \end{aligned}$$

If we replace $K_M(c_\lambda(\rho))$ and $m_\lambda(\rho)$ (from (5.10)) by their Taylor expansions of order up to n and recall the definition of $\beta_m(\lambda)$ from (5.20), we easily find that all the powers of order $< n + 1$ cancel out. Hence, there exists constant $C_1 > 0$ such that

$$|\Delta_\lambda(r)| \leq C_1 r^{n+1}$$

for $0 < r \leq \delta$. It follows that

$$\begin{aligned} & \left| \tilde{U}(r) - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} - \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} r^{1+\sigma(\lambda)} \right| \\ & \leq \frac{1}{r^3} \int_0^r |K_M(c_\lambda(\rho))| \left| P(\rho) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} \rho^{m-1} - \omega \rho^{\sigma(\lambda)} \right| \rho^3 d\rho + |\Delta_\lambda(r)| \\ & \leq \left(\frac{C_2 M}{n+4} + C_1 \right) r^{n+1} \end{aligned}$$

for $0 < r \leq \delta$, where C_2 denotes the maximum of $|K_M(c_\lambda(r))|$ for $0 \leq r \leq \delta$. Taking M_2 to be the constant on the right-hand side of the last inequality, we see that $\tilde{U}(r)$ satisfies (5.51). Thus $(\tilde{P}, \tilde{U}) \in X_\delta$ and \mathcal{F} maps \tilde{X}_δ into X_δ .

Next, we take any $(P_1, U_1), (P_2, U_2)$ in X_δ and set

$$(\tilde{P}_i, \tilde{U}_i) = \mathcal{F}(P_i, U_i), \quad i = 1, 2.$$

It is easy to verify that for any chosen $\theta \in (0, 1)$,

$$\begin{aligned} \sup_{0 < r \leq \delta} \left| f_\lambda(r, P_1(r), U_1(r)) - f_\lambda(r, P_2(r), U_2(r)) \right| / r^{1+\sigma(\lambda)-\theta} \\ \leq Cd((P_1, U_1), (P_2, U_2)). \end{aligned}$$

This enables us to deduce that

$$\sup_{0 < r \leq \delta} \left| \tilde{P}_1(r) - \tilde{P}_2(r) \right| / r^{\sigma(\lambda)} \leq C\delta^{2-\theta} d((P_1, U_1), (P_2, U_2)).$$

It is also readily seen that

$$\sup_{0 < r \leq \delta} \left| \tilde{U}_1(r) - \tilde{U}_2(r) \right| / r^{1+\sigma(\lambda)-\theta} \leq C\delta^\theta d((P_1, U_1), (P_2, U_2)),$$

so that

$$d((\tilde{P}_1, \tilde{U}_1), (\tilde{P}_2, \tilde{U}_2)) \leq C\delta^\theta d((P_1, U_1), (P_2, U_2)).$$

Thus \mathcal{F} is a contraction mapping in X_δ , provided δ is sufficiently small, and consequently, \mathcal{F} has a unique fixed point in X_δ .

The proof in the case $\gamma_{n+1}(\lambda) = 0$ ($\iff \sigma(\lambda) = n$) is similar. \square

Remark 5.1. The analytic solutions constructed in §4 for $\lambda \in (0, 1) \setminus S$ coincide with the solutions (5.43), (5.44) corresponding to $\omega = 0$.

From Theorem 5.3 we see that in the case $\gamma_1(\lambda) \leq 0$, the solution (p, u) of (2.19)–(2.21) satisfying $p, u \in C^1[0, \delta]$ and $p'(0) = 0$ is unique. If we weaken these regularity conditions, we can also get a 1-parameter family of solutions, as in the case $\gamma_1(\lambda) > 0$ and $\beta(\lambda) < 0$. More precisely, we have the following result:

Theorem 5.4. (1) Suppose that $\gamma_1(\lambda) = 0$ ($\iff \sigma(\lambda) = 0$). Then for any $\omega \in \mathbf{R}$, the problem (2.19)–(2.21) has a unique solution $(p(r), u(r))$ for $0 \leq r \leq \delta$, for some small $\delta > 0$, satisfying

$$(5.54) \quad p, u \in C^1[0, \delta] \quad \text{and} \quad p'(0) = \omega.$$

(2) Suppose that $\gamma_1(\lambda) < 0$ ($\iff -1 < \sigma(\lambda) < 0$). Then for any $\omega \in \mathbf{R}$, the problem (2.19)–(2.21) has a unique solution $(p(r), u(r))$ for $0 \leq r \leq \delta$, for some $\delta > 0$, satisfying the following conditions:

$$(5.55) \quad p \in C[0, \delta] \cap C^1(0, \delta], \quad u \in C^1[0, \delta];$$

$$p(r) = \alpha(\lambda) + \omega r^{1+\sigma(\lambda)} + O(r) \quad \text{as } r \rightarrow 0,$$

$$(5.56) \quad u(r) = \beta(\lambda)r + \frac{\omega K_M(\lambda)}{4 + \sigma(\lambda)} r^{2+\sigma(\lambda)} + O(r^2) \quad \text{as } r \rightarrow 0.$$

Proof. Consider the change of variables

$$(5.57) \quad p(r) = \alpha(\lambda) + r^{1+\sigma(\lambda)} P(r), \quad u(r) = r(\beta(\lambda) + U(r)),$$

where $P, U \in C[0, \delta] \cap C^1(0, \delta]$ and they satisfy the initial conditions

$$(5.58) \quad P(0) = \omega, \quad U(0) = 0.$$

One can readily verify that, under the transformation (5.57), the equations (2.19) and (2.21) are respectively changed into the equations

$$(5.59) \quad U'(r) + \frac{3}{r}U(r) = r^{\sigma(\lambda)} K_M(c_\lambda(r)) P(r) + m_\lambda(r),$$

$$(5.60) \quad P'(r) = \tilde{f}_\lambda(r, P(r), U(r)),$$

where $m_\lambda(r)$ is as in (5.10), and $\tilde{f}_\lambda(r, P, U)$ is given by

$$\begin{aligned} \tilde{f}_\lambda(r, P, U) = & \left(-r^{\sigma(\lambda)} K_M(c_\lambda(r)) P^2 + \frac{g_\lambda(0)}{\beta(\lambda)} \frac{UP}{r} \right. \\ & \left. - \frac{g_\lambda(r) - g_\lambda(0)}{r} P + \frac{h_\lambda(r)}{r^{2+\sigma(\lambda)}} \right) / (\beta(\lambda) + U), \end{aligned}$$

where $g_\lambda(r)$, $h_\lambda(r)$ are as in (5.12) and (5.13), respectively. Since $-1 < \sigma(\lambda) \leq 0$, the term $r^{\sigma(\lambda)} K_M(c_\lambda(r))$ in (5.59) and \tilde{f}_λ are integrable. Also, the term UP/r is integrable if $U(r)$ converges to 0 at an appropriate rate as $r \rightarrow 0$. This enables us to transform the problem (5.58)–(5.60) into the following system of integral equations:

$$(5.61) \quad U(r) = \frac{1}{r^3} \int_0^r K_M(c_\lambda(\rho)) P(\rho) \rho^{3+\sigma(\lambda)} d\rho + \frac{1}{r^3} \int_0^r m_\lambda(\rho) \rho^3 d\rho,$$

$$(5.62) \quad P(r) = \omega + \int_0^r \tilde{f}_\lambda(\rho, P(\rho), U(\rho)) d\rho.$$

Using the contraction mapping principle as in the proof of Theorem 5.3, we can prove that there exists $\delta > 0$ such that the system (5.61), (5.62) has a unique continuous solution $(P(r), U(r))$ for $0 \leq r \leq \delta$, satisfying

$$|P(r)| \leq M_1, \quad |U(r)| \leq M_2 r^{1+\sigma(\lambda)} \quad (0 \leq r \leq \delta)$$

for some constants M_1 and M_2 . Hence the desired result follows. □

Later in §8 we shall see that the solutions of (2.19)–(2.21) constructed in this section are unique also within the class of nonnegative weak solutions (see Theorem 8.1).

6. CONTINUOUS DEPENDENCE OF SOLUTIONS ON THE PARAMETERS

By the results obtained in the previous section, for every $0 < \lambda < \lambda_\infty$ the problem (2.19)–(2.21) has a continuum of solutions depending on a real parameter ω . From (5.23)–(5.27) and (5.62) we see, at least heuristically, that, for a fixed ω , the solution depends continuously on λ for $0 < \lambda < \lambda_\infty$, $\lambda \notin S_1$, where

$$\begin{aligned} S_1 &= \{ \lambda \in (0, \lambda_\infty) : \gamma_n(\lambda) = 0 \text{ for some } n \geq 2 \} \\ &= \{ \lambda \in (0, \lambda_\infty) : \sigma(\lambda) = m \text{ for some integer } m \geq 1 \}. \end{aligned}$$

However, from the structure of the solutions (in Theorem 5.3) it is clear that as $\lambda \rightarrow \bar{\lambda}$ for some $\bar{\lambda} \in S_1$, the solution does not have a limit. The main purpose of this section is to show that we can reparameterize the solutions for $0 < \lambda < \lambda_\infty$ by a new parameter ψ in such a way that when $\omega = \omega(\lambda, \psi)$ the solution will depend continuously on λ for all $0 < \lambda < \lambda_\infty$, as well as on ψ .

We introduce the set

$$A = \{ \lambda \in (0, \lambda_\infty) : \gamma_1(\lambda) > 0 \} = \{ \lambda \in (0, \lambda_\infty) : \sigma(\lambda) > 0 \},$$

which is clearly an open set containing S_1 . By Lemma 3.3, the following two situations may occur: (1) A is a single interval $(\lambda_0, \lambda_\infty)$, where λ_0 is either equal to zero (in the case $\sigma(\lambda) > 0$ for all $0 < \lambda < \lambda_\infty$) or equal to the unique positive root of the equation $\sigma(\lambda) = 0$ (in the case that $\sigma(\lambda)$ changes sign only once); (2) A is a union of finitely many disjoint open intervals, one of which has the form $(\lambda_0, \lambda_\infty)$, where λ_0 is the largest root of the equation $\sigma(\lambda) = 0$. For any $\lambda \in A$ and $\omega \in \mathbf{R}$, we denote by $(P(r; \lambda, \omega), U(r; \lambda, \omega))$ the solution of the system consisting of (5.17) and

one of the equations (5.23)–(5.27) (in accordance with the corresponding condition on λ), and for any $\lambda \in (0, \lambda_\infty) \setminus A$ we denote by $(P(r; \lambda, \omega), U(r; \lambda, \omega))$ the solution of the system of equations (5.61) and (5.62).

Consider Case (1) first. By Lemma 3.3, the set S_1 consists of a finite number of points $\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_m$, which are the roots of the equations $\gamma_2(\lambda) = 0$, $\gamma_3(\lambda) = 0$, \dots , $\gamma_{n_0-1}(\lambda) = 0$, and the sequence $\{\lambda_n\}_{n \geq n_0}$ as described by Lemma 3.3. We rearrange the union of the two sets of these λ 's as a monotone increasing sequence and denote it as

$$(6.1) \quad \bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_n, \dots$$

By Lemma 3.3 we know that $\lim_{n \rightarrow \infty} \bar{\lambda}_n = \lambda_\infty$. For each integer $n \geq 1$ we take a small number $\varepsilon_n > 0$ such that all the intervals

$$[\bar{\lambda}_1 - \varepsilon_1, \bar{\lambda}_1 + \varepsilon_1], [\bar{\lambda}_2 - \varepsilon_2, \bar{\lambda}_2 + \varepsilon_2], \dots, [\bar{\lambda}_n - \varepsilon_n, \bar{\lambda}_n + \varepsilon_n], \dots$$

are contained in $(\lambda_0, \lambda_\infty)$ and do not intersect each other. Let $m_n = \sigma(\bar{\lambda}_n)$. Clearly, if $\sigma(\lambda)$ is monotone increasing, then $m_{n+1} = m_n + 1$; for general $\sigma(\lambda)$ we only have $|m_{n+1} - m_n| \leq 1$. We introduce, for every integer $n \geq 1$, a function $\omega_n(\lambda, \psi)$ ($\lambda \in [\bar{\lambda}_n - \varepsilon_n, \bar{\lambda}_n + \varepsilon_n]$, $\psi \in \mathbf{R}$) as follows:

$$(6.2) \quad \omega_n(\lambda, \psi) = \begin{cases} \psi & \text{if } \lambda = \bar{\lambda}_n, \\ \psi - (\mu_{m_n-1}(\lambda)/(m_n - \sigma(\lambda))) & \text{if } 0 < |\lambda - \bar{\lambda}_n| \leq \varepsilon_n. \end{cases}$$

We arbitrarily take a continuous function $\omega_{n,n+1}(\lambda, \psi)$ defined for $\bar{\lambda}_n + \varepsilon_n \leq \lambda \leq \bar{\lambda}_{n+1} - \varepsilon_{n+1}$ and $\psi \in \mathbf{R}$ and monotone increasing in ψ such that

$$(6.3) \quad \begin{aligned} \omega_{n,n+1}(\bar{\lambda}_n + \varepsilon_n, \psi) &= \omega_n(\bar{\lambda}_n + \varepsilon_n, \psi), \\ \omega_{n,n+1}(\bar{\lambda}_{n+1} - \varepsilon_{n+1}, \psi) &= \omega_{n+1}(\bar{\lambda}_{n+1} - \varepsilon_{n+1}, \psi). \end{aligned}$$

For $\lambda_0 < \lambda \leq \bar{\lambda}_1 - \varepsilon_1$ we take

$$(6.4) \quad \omega_{0,1}(\lambda, \psi) = \psi - (\mu_{n_1-1}(\lambda)/(n_1 - \sigma(\lambda))).$$

Finally, in case $\lambda_0 > 0$ we define

$$(6.5) \quad \omega_0(\lambda, \psi) = \psi - \mu_0(\lambda) \quad (0 < \lambda \leq \lambda_0, \psi \in \mathbf{R}).$$

Now define, for all $0 < \lambda < \lambda_\infty$ and $\psi \in \mathbf{R}$, a function $\omega(\lambda, \psi)$ by

$$(6.6) \quad \omega(\lambda, \psi) = \begin{cases} \omega_0(\lambda, \psi) & \text{if } 0 < \lambda \leq \lambda_0, \\ \omega_{0,1}(\lambda, \psi) & \text{if } \lambda_0 < \lambda \leq \bar{\lambda}_1 - \varepsilon_1, \\ \omega_n(\lambda, \psi) & \text{if } \bar{\lambda}_n - \varepsilon_n < \lambda < \bar{\lambda}_n + \varepsilon_n, \quad n = 1, 2, \dots, \\ \omega_{n,n+1}(\lambda, \psi) & \text{if } \bar{\lambda}_n + \varepsilon_n \leq \lambda \leq \bar{\lambda}_{n+1} - \varepsilon_{n+1}, \quad n = 1, 2, \dots \end{cases}$$

It follows that $\omega(\lambda, \psi)$ is continuous for $\lambda \in (0, \lambda_\infty) \setminus S_1$ and $\psi \in \mathbf{R}$, and is monotone increasing in ψ . Notice however that $\omega(\lambda, \psi)$ is not continuous at $\bar{\lambda} \in S_1$, since

$$\lim_{\lambda \rightarrow \bar{\lambda}} \omega(\lambda, \psi) = \pm\infty.$$

Consider Case (2) next. We write

$$A = (\lambda_{01}, \lambda_{02}) \cup (\lambda_{03}, \lambda_{04}) \cup \dots \cup (\lambda_{0,2k-1}, \lambda_{0,2k}) \cup (\lambda_0, \lambda_\infty).$$

In the last interval we can define the function $\omega(\lambda, \psi)$ in a way similar to (6.6). Every interval of the form $(\lambda_{0,2i-1}, \lambda_{0,2i})$ contains at most a finite number of points from S_1 . We write these points in increasing order, say,

$$\bar{\lambda}_{i1}, \bar{\lambda}_{i2}, \dots, \bar{\lambda}_{im}.$$

As before, for every integer $1 \leq j \leq m$ we take a small positive number ε_j such that all the intervals

$$[\bar{\lambda}_{i1} - \varepsilon_1, \bar{\lambda}_{i1} + \varepsilon_1], [\bar{\lambda}_{i2} - \varepsilon_2, \bar{\lambda}_{i2} + \varepsilon_2], \dots, [\bar{\lambda}_{im} - \varepsilon_m, \bar{\lambda}_{im} + \varepsilon_m]$$

are contained in $(\lambda_{0\,2i-1}, \lambda_{0\,2i})$ and do not intersect each other. We then define $\omega(\lambda, \psi)$ for $\lambda \in [\bar{\lambda}_{ij} - \varepsilon_j, \bar{\lambda}_{ij} + \varepsilon_j]$ and $\psi \in \mathbf{R}$ ($j = 1, 2, \dots, m$) similarly as in (6.2), and, for every $1 \leq j \leq m - 1$, we arbitrarily define the values of $\omega(\lambda, \psi)$ for $\lambda \in (\bar{\lambda}_{ij} + \varepsilon_j, \bar{\lambda}_{i,j+1} - \varepsilon_{j+1})$ and $\psi \in \mathbf{R}$ such that it is continuous on

$$[\bar{\lambda}_{ij} + \varepsilon_j, \bar{\lambda}_{i,j+1} - \varepsilon_{j+1}] \times \mathbf{R}$$

and monotone increasing in ψ . For $\lambda \in (\lambda_{0\,2i-1}, \bar{\lambda}_{i1} - \varepsilon_1) \cup (\bar{\lambda}_{im} + \varepsilon_m, \lambda_{0\,2i})$, we define $\omega(\lambda, \psi)$ similarly as in (6.4). Finally, in the complementary set of A , which consists of the union $(0, \lambda_{01}] \cup [\lambda_{02}, \lambda_{03}] \cup [\lambda_{04}, \lambda_{05}] \cup \dots \cup [\lambda_{0\,2k}, \lambda_0]$, we define

$$\omega(\lambda, \psi) = \psi - \mu_0(\lambda) \quad \text{for } \psi \in \mathbf{R}.$$

Clearly, the function $\omega(\lambda, \psi)$ ($\lambda \in (0, \lambda_\infty)$, $\psi \in \mathbf{R}$) defined in this way is continuous for all $\lambda \in (0, \lambda_\infty) \setminus S_1$ and $\psi \in \mathbf{R}$, but not for $\lambda \in S_1$, and is monotone increasing in ψ .

Having introduced the function $\omega(\lambda, \psi)$ ($0 < \lambda < \lambda_\infty$, $\psi \in \mathbf{R}$), we now define

$$(6.7) \quad P_{\lambda\psi}(r) = P(r; \lambda, \omega(\lambda, \psi)), \quad U_{\lambda\psi}(r) = U(r; \lambda, \omega(\lambda, \psi))$$

and

$$(6.8) \quad p_{\lambda\psi}(r) = \alpha(\lambda) + rP_{\lambda\psi}(r), \quad u_{\lambda\psi}(r) = r(\beta(\lambda) + U_{\lambda\psi}(r))$$

for all $0 < \lambda < \lambda_\infty$ and $\psi \in \mathbf{R}$.

The main result of this section is the following:

Theorem 6.1. *For any $0 < \bar{\lambda} < \lambda_\infty$ and $\bar{\psi} \in \mathbf{R}$, there exists a corresponding $\delta > 0$ such that for any (λ, ψ) in a neighborhood of $(\bar{\lambda}, \bar{\psi})$, the solution $(p_{\lambda\psi}(r), u_{\lambda\psi}(r))$ exists for all $r \in [0, \delta]$, and*

$$(6.9) \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} p_{\lambda\psi}(r) = p_{\bar{\lambda}\bar{\psi}}(r), \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} u_{\lambda\psi}(r) = u_{\bar{\lambda}\bar{\psi}}(r)$$

uniformly for $r \in [0, \delta]$.

We shall only give the proof for the case $(\bar{\lambda}, \bar{\psi}) \in S_1 \times \mathbf{R}$; the proof for the case $\bar{\lambda} \in (0, \lambda_\infty) \setminus S_1$ (in particular, the case where $\sigma(\bar{\lambda}) = 0$) is similar but simpler. We need the following lemma:

Lemma 6.2 *Let $\bar{\lambda} \in S_1$, $\bar{\psi} \in \mathbf{R}$. Then there exist positive constants δ, C such that for $0 < |\lambda - \bar{\lambda}| \ll 1$ and $|\psi - \bar{\psi}| \ll 1$, the solution $(P_{\lambda\psi}(r), U_{\lambda\psi}(r))$ exists for all $0 \leq r \leq \delta$ and*

$$(6.10) \quad \left| P_{\lambda\psi}(r) - \psi r^{\sigma(\lambda)} - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} - \mu_{n-1}(\lambda) \frac{r^n - r^{\sigma(\lambda)}}{n - \sigma(\lambda)} \right| \leq Cr^{n+1},$$

where $n = \sigma(\bar{\lambda})$; the constants δ, C are independent of λ and ψ .

Proof. Notice that since $\bar{\lambda} \in S_1$, $n = \sigma(\bar{\lambda})$ is an integer ≥ 1 . In the following, when we say “ λ near $\bar{\lambda}$ ”, we mean that $0 < |\lambda - \bar{\lambda}| \ll 1$, and similarly for “ ψ near $\bar{\psi}$ ”.

Since $P_{\lambda\psi}(r) = P(r; \lambda, \omega(\lambda, \psi))$ and $\omega(\lambda, \psi) = \psi - \mu_{n-1}(\lambda)/(n - \sigma(\lambda))$ for λ near $\bar{\lambda}$, we see that $P_{\lambda\psi}(r)$ satisfies the equation

$$(6.11) \quad \begin{aligned} P(r) = & \psi r^{\sigma(\lambda)} + \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} + \mu_{n-1}(\lambda) \frac{r^n - r^{\sigma(\lambda)}}{n - \sigma(\lambda)} \\ & + r^{\sigma(\lambda)} \int_0^r \left(f_{\lambda}(\rho, P(\rho), U(\rho)) - \sum_{m=0}^{n-1} \mu_m(\lambda) \rho^m \right) \rho^{-\sigma(\lambda)} d\rho. \end{aligned}$$

As in the proof of Theorem 5.3, we then have

$$(6.12) \quad P(r) = \omega r^{\sigma(\lambda)} + \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} + \mu_{n-1}(\lambda) \frac{r^n - r^{\sigma(\lambda)}}{n - \sigma(\lambda)} + O(r^{n'})$$

as $r \rightarrow 0$, where $n' = n$ for $n - 1 < \sigma(\lambda) < n$, and $n' = n + 1$ for $n < \sigma(\lambda) < n + 1$. Thus for λ near $\bar{\lambda}$ and for all $\psi \in \mathbf{R}$, the function

$$(6.13) \quad \Delta_{\lambda\psi}(r) = \sup_{0 < \rho \leq r} \left| P_{\lambda\psi}(\rho) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} \rho^{m-1} \right| \rho^{-n+\frac{1}{2}}$$

converges to zero as $r \rightarrow 0$. Let $\delta_0 > 0$ be a fixed small number, and let $\delta_{\lambda\psi}$ be the largest number in $(0, \delta_0]$ such that the solution of the system (6.11), (5.17) exists for $0 \leq r < \delta_{\lambda\psi}$. We can clearly write, for $0 < \rho \leq r < \delta_{\lambda\psi}$,

$$(6.14) \quad \left| P_{\lambda\psi}(\rho) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} \rho^{m-1} \right| \leq \rho^{n-\frac{1}{2}} \Delta_{\lambda\psi}(r).$$

By (5.17) we further have

$$(6.15) \quad \begin{aligned} \left| U_{\lambda\psi}(\rho) - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} \rho^{m-1} \right| & \leq C \rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}(\rho) + C \rho^{n+1} \\ & \leq C \rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}(r) + C \rho^{n+1} \end{aligned}$$

for $0 < \rho \leq r < \delta_{\lambda\psi}$; here and in what follows we use C to denote various positive constants *independent of* λ and ψ . We introduce the functions

$$v_{\lambda\psi}(r) = P_{\lambda\psi}(r) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1}, \quad w_{\lambda\psi}(r) = U_{\lambda\psi}(r) - \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1},$$

so that

$$(6.16) \quad \begin{aligned} P_{\lambda\psi}(r) &= \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} + v_{\lambda\psi}(r), \\ U_{\lambda\psi}(r) + \beta(\lambda) &= \sum_{m=1}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} + w_{\lambda\psi}(r) \end{aligned}$$

(recall that $\beta_1(\lambda) = \beta(\lambda)$, $\beta_2(\lambda) = 0$). By (6.14), (6.15),

$$(6.17) \quad |v_{\lambda\psi}(\rho)| \leq \rho^{n-\frac{1}{2}} \Delta_{\lambda\psi}(r), \quad |w_{\lambda\psi}(\rho)| \leq C \rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}(r) + C \rho^{n+1}$$

for $0 < \rho \leq r < \delta_{\lambda\psi}$. It follows (see the Appendix) that

$$(6.18) \quad \begin{aligned} & (U_{\lambda\psi}(\rho) + \beta(\lambda)) \left(f_{\lambda}(\rho, P_{\lambda\psi}(\rho), U_{\lambda\psi}(\rho)) - \sum_{m=0}^{n-1} \mu_m(\lambda) \rho^m \right) \\ &= \rho^n y_{\lambda\psi}(\rho) + z_{1\lambda\psi}(\rho) w_{\lambda\psi}(\rho) + \rho z_{2\lambda\psi}(\rho) v_{\lambda\psi}(\rho) - K_M(c_{\lambda}(\rho)) v_{\lambda\psi}^2(\rho), \end{aligned}$$

where

$$(6.19) \quad |y_{\lambda\psi}(\rho)| \leq C, \quad |z_{1\lambda\psi}(\rho)| \leq C, \quad |z_{2\lambda\psi}(\rho)| \leq C$$

for $0 < \rho < \delta_{\lambda\psi}$ and for λ, ψ respectively near $\bar{\lambda}$ and $\bar{\psi}$. Hence,

$$(6.20) \quad \begin{aligned} & |U_{\lambda\psi}(\rho) + \beta(\lambda)| \left| f_{\lambda}(\rho, P_{\lambda\psi}(\rho), U_{\lambda\psi}(\rho)) - \sum_{m=0}^{n-1} \mu_m(\lambda) \rho^m \right| \\ & \leq C \rho^{2n-1} \Delta_{\lambda\psi}^2(r) + C \rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}(r) + C \rho^n. \end{aligned}$$

By the second inequality in (6.18) it follows that

$$\begin{aligned} |U_{\lambda\psi}(\rho) + \beta(\lambda)| & \geq |\beta(\lambda)| - |U_{\lambda\psi}(\rho)| \\ & \geq |\beta(\lambda)| - (Cr^2 + Cr^{n+\frac{1}{2}} \Delta_{\lambda\psi}(r)). \end{aligned}$$

Since $\lim_{r \rightarrow 0} \Delta_{\lambda\psi}(r) = 0$ and $\beta(\lambda) \neq 0$, we infer that for every (λ, ψ) near $(\bar{\lambda}, \bar{\psi})$, there exists a corresponding positive number $\tilde{\delta}_{\lambda\psi} \leq \delta_{\lambda\psi}$ such that for $0 < r < \tilde{\delta}_{\lambda\psi}$,

$$(6.21) \quad Cr^2 + Cr^{\frac{3}{2}} \Delta_{\lambda\psi}(r) \leq \frac{1}{2} |\beta(\lambda)|.$$

Thus

$$(6.22) \quad |U_{\lambda\psi}(\rho) + \beta(\lambda)| \geq \frac{1}{2} |\beta(\lambda)| \geq C > 0$$

for $0 < \rho < \tilde{\delta}_{\lambda\psi}$ and for λ, ψ respectively near $\bar{\lambda}$ and $\bar{\psi}$. In what follows we use the same notation $\tilde{\delta}_{\lambda\psi}$ to denote the supremum of all $\tilde{\delta}_{\lambda\psi}$ such that (6.21) holds for $0 < r < \tilde{\delta}_{\lambda\psi}$. Substituting (6.22) into (6.20), we get

$$\begin{aligned} & \left| f_{\lambda}(\rho, P_{\lambda\psi}(\rho), U_{\lambda\psi}(\rho)) - \sum_{m=0}^{n-1} \mu_m(\lambda) \rho^m \right| \\ & \leq C \rho^{2n-1} \Delta_{\lambda\psi}^2(r) + C \rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}(r) + C \rho^n \end{aligned}$$

and then, by (6.11),

$$(6.23) \quad \begin{aligned} & \left| P_{\lambda\psi}(r) - \psi r^{\sigma(\lambda)} - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} - \mu_{n-1}(\lambda) \frac{r^n - r^{\sigma(\lambda)}}{n - \sigma(\lambda)} \right| \\ & \leq Cr^{2n} \Delta_{\lambda\psi}^2(r) + Cr^{n+\frac{3}{2}} \Delta_{\lambda\psi}(r) + Cr^{n+1} \end{aligned}$$

for $0 < r < \tilde{\delta}_{\lambda\psi}$. The term with the coefficient $\mu_{n-1}(\lambda)$ can be estimated from the inequality

$$\left| \frac{r^n - r^{\sigma}}{n - \sigma} \right| \leq Cr^{\frac{1}{2}}$$

(for $r > 0$, $\sigma \geq n - 1/4$), and this leads to

$$\begin{aligned}
 (6.24) \quad & \left| P_{\lambda\psi}(\rho) - \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} \rho^{m-1} \right| \rho^{-n+\frac{1}{2}} \\
 & \leq C\rho^{n+\frac{1}{2}} \Delta_{\lambda\psi}^2(\rho) + C\rho^2 \Delta_{\lambda\psi}(\rho) + C\rho^{\frac{3}{2}} + C\rho^{\frac{1}{4}} \\
 & \leq C\rho^{\frac{3}{2}} \Delta_{\lambda\psi}^2(r) + C\rho^{\frac{1}{4}}
 \end{aligned}$$

for $0 < \rho \leq r < \tilde{\delta}_{\lambda\psi}$, where λ is sufficiently near $\bar{\lambda}$ such that $|\sigma(\lambda) - n| < 1/4$, and ψ is near $\bar{\psi}$. By (6.13), it then follows that, for $0 < r < \tilde{\delta}_{\lambda\psi}$ and for λ, ψ respectively near $\bar{\lambda}$ and $\bar{\psi}$,

$$\Delta_{\lambda\psi}(r) \leq Cr^{\frac{3}{2}} \Delta_{\lambda\psi}^2(r) + Cr^{\frac{1}{4}}.$$

Thus, if $0 < r < \tilde{\delta}_{\lambda\psi}$ and

$$(6.25) \quad 4Cr^{\frac{3}{2}} \cdot Cr^{\frac{1}{4}} < 1,$$

then

$$(6.26) \quad \Delta_{\lambda\psi}(r) \leq \frac{1}{2Cr^{\frac{3}{2}}} \left(1 - \sqrt{1 - 4Cr^{\frac{3}{2}} \cdot Cr^{\frac{1}{4}}} \right) < 2Cr^{\frac{1}{4}}.$$

We now replace the term $\Delta_{\lambda\psi}(r)$ in (6.21) by the upper bound obtained in (6.26) and consider the inequality

$$(6.27) \quad Cr^2 + Cr^{\frac{3}{2}} \cdot 2Cr^{\frac{1}{4}} \leq \frac{1}{2} |\beta(\lambda)|.$$

Since $|\beta(\lambda)|$ has a positive lower bound for λ near $\bar{\lambda}$ and the left-hand side converges to zero as $r \rightarrow 0$, we can find $\delta \in (0, \delta_0)$ small enough so that (6.27) holds for all $0 < r \leq \delta$ and λ near $\bar{\lambda}$. By taking δ sufficiently small we may assume that (6.25) also holds for $0 < r \leq \delta$. We now use the maximality of $\tilde{\delta}_{\lambda\psi}$ and $\delta_{\lambda\psi}$ to prove that

$$\delta \leq \tilde{\delta}_{\lambda\psi} (\leq \delta_{\lambda\psi})$$

for λ, ψ near $\bar{\lambda}$ and $\bar{\psi}$, respectively. Indeed, we have either $\tilde{\delta}_{\lambda\psi} = \delta_{\lambda\psi}$ or $\tilde{\delta}_{\lambda\psi} < \delta_{\lambda\psi}$. Suppose first that $\tilde{\delta}_{\lambda\psi} = \delta_{\lambda\psi}$. If $\delta_{\lambda\psi} < \delta$, then, by (6.22), (6.23) and (6.26), $P_{\lambda\psi}(r)$ and, consequently, $U_{\lambda\psi}(r)$ are bounded for $0 \leq r \leq \delta_{\lambda\psi}$, and $U_{\lambda\psi}(r) + \beta(\lambda)$ stays away from zero uniformly for $0 \leq r \leq \delta_{\lambda\psi}$. This allows us to extend $(P_{\lambda\psi}(r), U_{\lambda\psi}(r))$ beyond $r = \delta_{\lambda\psi}$ as the solution of a regular ODE system, which contradicts the maximality of $\delta_{\lambda\psi}$. Hence $\delta \leq \delta_{\lambda\psi}$ holds in the case where $\tilde{\delta}_{\lambda\psi} = \delta_{\lambda\psi}$. Suppose next that $\tilde{\delta}_{\lambda\psi} < \delta_{\lambda\psi}$. Then by the maximality of $\tilde{\delta}_{\lambda\psi}$, the equality in (6.21) must hold for $r = \tilde{\delta}_{\lambda\psi}$. Since the second inequality in (6.26) holds also for $r = \min\{\delta, \tilde{\delta}_{\lambda\psi}\}$, it follows, by (6.27), that $\delta < \tilde{\delta}_{\lambda\psi}$. Having thus completed the proof that $\delta \leq \delta_{\lambda\psi}$, the estimate (6.10) follows from (6.23) and (6.26). \square

Proof of Theorem 6.1. We shall only consider the case $\bar{\lambda} \in S_1$; the case $\bar{\lambda} \in (0, \lambda_\infty) \setminus S_1$ can be treated similarly.

We first prove that for δ as in Lemma 6.2,

$$(6.28) \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} P_{\lambda\psi}(r) = P_{\bar{\lambda}\bar{\psi}}(r), \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} U_{\lambda\psi}(r) = U_{\bar{\lambda}\bar{\psi}}(r)$$

uniformly for all $r \in [0, \delta]$.

From Lemma 6.2 we see that

$$(6.29) \quad |P_{\lambda\psi}(r)| \leq Cr^{\frac{3}{4}}$$

for $0 \leq r \leq \delta$ and (λ, ψ) near $(\bar{\lambda}, \bar{\psi})$, which further implies, by (5.17), that

$$(6.30) \quad |U_{\lambda\psi}(r)| \leq Cr^{\frac{7}{4}}$$

for $0 \leq r \leq \delta$ and (λ, ψ) near $(\bar{\lambda}, \bar{\psi})$. It follows that

$$(6.31) \quad |f_{\lambda}(r, P_{\lambda\psi}(r), U_{\lambda\psi}(r))| \leq C$$

for $0 \leq r \leq \delta$ and (λ, ψ) near $(\bar{\lambda}, \bar{\psi})$. Using these estimates in (5.7), (5.8), we obtain

$$(6.32) \quad |U'_{\lambda\psi}(r)| \leq C,$$

$$(6.33) \quad |P'_{\lambda\psi}(r)| \leq Cr^{-\frac{1}{4}} + C.$$

By (6.32), (6.33) and (6.29), it follows that the family of functions

$$\{(U_{\lambda\psi}(r), P_{\lambda\psi}(r)) : 0 < |\lambda - \bar{\lambda}| \ll 1, |\psi - \bar{\psi}| \ll 1\}$$

is equicontinuous and uniformly bounded for $0 \leq r \leq \delta$. We now assert that

$$(6.34) \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} \max_{0 \leq r \leq \delta} |P_{\lambda\psi}(r) - P_{\bar{\lambda}\bar{\psi}}(r)| = 0, \quad \lim_{(\lambda, \psi) \rightarrow (\bar{\lambda}, \bar{\psi})} \max_{0 \leq r \leq \delta} |U_{\lambda\psi}(r) - U_{\bar{\lambda}\bar{\psi}}(r)| = 0.$$

Indeed, otherwise there would exist a number $\varepsilon > 0$ and a sequence $(\bar{\lambda}_m, \bar{\psi}_m)$ ($m = 1, 2, \dots$), converging to $(\bar{\lambda}, \bar{\psi})$, such that

$$(6.35) \quad \max_{0 \leq r \leq \delta} |P_{\bar{\lambda}_m \bar{\psi}_m}(r) - P_{\bar{\lambda}\bar{\psi}}(r)| + \max_{0 \leq r \leq \delta} |U_{\bar{\lambda}_m \bar{\psi}_m}(r) - U_{\bar{\lambda}\bar{\psi}}(r)| \geq \varepsilon$$

for all m . Since $\{P_{\bar{\lambda}_m \bar{\psi}_m}(r)\}$ and $\{U_{\bar{\lambda}_m \bar{\psi}_m}(r)\}$ are both equicontinuous and uniformly bounded for $0 \leq r \leq \varepsilon$, there exists a subsequence of $\{(\bar{\lambda}_m, \bar{\psi}_m)\}$, which for simplicity is again denoted by $\{(\bar{\lambda}_m, \bar{\psi}_m)\}$, such that $P_{\bar{\lambda}_m \bar{\psi}_m}(r)$ and $U_{\bar{\lambda}_m \bar{\psi}_m}(r)$ converge uniformly, in $[0, \delta]$, to some functions $P(r)$ and $U(r)$, respectively. Taking $(\lambda, \psi) = (\bar{\lambda}_m, \bar{\psi}_m)$ in (5.17), (6.11) and using the fact that

$$\lim_{\sigma \rightarrow n} (r^n - r^\sigma) / (n - \sigma) = r^n \log r$$

and the Lebesgue dominated convergence theorem, we conclude, as $m \rightarrow \infty$, that $P(r)$, $U(r)$ form a solution for the system of (5.17), (5.27). Since the solution of the system (5.17), (5.27) is unique, we get

$$P(r) = P_{\bar{\lambda}\bar{\psi}}(r), \quad U(r) = U_{\bar{\lambda}\bar{\psi}}(r),$$

a contradiction to (6.35). Hence (6.34) holds. By (6.34) and (6.8), (6.9) easily follows. \square

7. GLOBAL PROPERTIES OF THE SOLUTIONS

In this section we study the global properties of solutions of the free boundary problem (2.19)–(2.22) and the initial value problem (2.19)–(2.21).

We first consider the global properties of solutions of the free boundary problem (2.19)–(2.22).

Theorem 7.1. *Let (p, u, R_λ) be a solution of the free boundary problem (2.19)–(2.22) with $p(r)$, $u(r)$ in $C^1[0, R_\lambda]$. Then*

- (1) $p'(r) > 0$ for $0 < r < R_\lambda$;
- (2) $0 < p(r) < 1$ for $0 \leq r < R_\lambda$, $p(R_\lambda) = 1$;
- (3) $\frac{d}{dr}(u'(r) + \frac{2}{r}u(r)) > 0$ for $0 < r < R_\lambda$;
- (4) $u(r) < 0$ for $0 < r < R_\lambda$;
- (5) $u'(0) < 0$, $u'(R_\lambda) > 0$.

Proof. First we note that the assumptions that $p, u \in C^1[0, R_\lambda]$ and $u(0) = 0$ imply, by (2.19), that $u \in C^2[0, R_\lambda]$.

We begin by proving that $p'(r) \geq 0$ for all $0 \leq r \leq R_\lambda$. Assume that this assertion is false, namely, there exists an $r_0 \in (0, R_\lambda)$ such that $p'(r_0) < 0$. Denote by (r_1, r_2) the maximal interval containing r_0 such that $p'(r) < 0$ for all $r \in (r_1, r_2)$. Then either $r_1 = 0$ or $0 < r_1 < r_0$ and $p'(r_1) = 0$. In both situations we get $u(r_1)p'(r_1) = 0$. This implies, by (2.21) and Lemma 3.1, that

$$p(r_1) = \alpha(c_\lambda(r_1)).$$

Similarly, we have either $r_2 = R_\lambda$, or $r_0 < r_2 < R_\lambda$ and $p'(r_2) = 0$. In both situations we get $u(r_2)p'(r_2) = 0$, so that

$$p(r_2) = \alpha(c_\lambda(r_2)).$$

Since $p'(r) < 0$ for $r \in (r_1, r_2)$, it follows that $p(r_1) > p(r_2)$. On the other hand, since $\alpha(\lambda)$ is monotone increasing in λ (by Lemma 3.1) and $c_\lambda(r)$ is monotone increasing in r , we have $\alpha(c_\lambda(r_1)) < \alpha(c_\lambda(r_2))$, which is a contradiction. Hence $p'(r) \geq 0$ for $0 \leq r \leq R_\lambda$. A similar argument shows that $p'(r)$ cannot be identically zero on any intervals, so that $p(r)$ is strictly monotone increasing.

The assertion (2) follows immediately from the above result, because by Lemma 3.1 we have $p(0) = \alpha(\lambda) > 0$ and $p(R_\lambda) = \alpha(1) = 1$. Differentiating the equation (2.19), we get

$$\begin{aligned} \frac{d}{dr}(u'(r) + \frac{2}{r}u(r)) &= -K'_D(c_\lambda(r))c'_\lambda(r) + K'_M(c_\lambda(r))c'_\lambda(r)p(r) + K'_N(c_\lambda(r))p'(r) \\ &> 0 \quad \text{for } 0 < r \leq R_\lambda, \end{aligned}$$

because $K'_D < 0$, $c'_\lambda > 0$, $K'_M > 0$, $p > 0$, $K'_N > 0$ and $p' \geq 0$, so that (3) holds.

Next we prove assertion (4). Assume that this assertion is false. Since clearly $u(r)$ cannot be identically zero, it follows that two situations may take place: either (a) $\max_{0 \leq r \leq R_\lambda} u(r) > 0$ or (b) $u(r) \leq 0$ for all $0 \leq r \leq R_\lambda$, $\min_{0 \leq r \leq R_\lambda} u(r) < 0$, and there exists an $r_0 \in (0, R_\lambda)$ such that $u(r_0) = 0$. In case (a) we let $u(\tilde{r}_0) = \max_{0 \leq r \leq R_\lambda} u(r)$. It follows that $u'(\tilde{r}_0) = 0$ and $u(\tilde{r}_0) > 0$, so that

$$u'(\tilde{r}_0) + \frac{2}{\tilde{r}_0}u(\tilde{r}_0) > 0.$$

On the other hand, denoting by (r_1, r_2) the maximal interval containing \tilde{r}_0 such that $u(r) > 0$ for all $r \in (r_1, r_2)$, we have $u(r_2) = 0$ and $u'(r_2) \leq 0$. It follows that

$$u'(r_2) + \frac{2}{r_2}u(r_2) \leq 0.$$

This contradicts assertion (3). Similarly, in case (b), by comparing the values of $u'(r) + (2/r)u(r)$ at the points r_0, r_2 , where $r_2 \in (r_0, R_\lambda)$ is such that $u(r_2) = \min_{r_0 \leq r \leq R_\lambda} u(r)$, we again get a contradiction. Hence assertion (4) follows.

We now prove that $p'(r) > 0$ for $0 < r < R_\lambda$. Indeed, since $p'(r) \geq 0$ and $u(r) < 0$, we have $u(r)p'(r) \leq 0$, so that, by equation (2.21),

$$K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda(r)))p(r) - K_M(c_\lambda(r))p^2(r) \leq 0$$

for $0 \leq r \leq R_\lambda$. It follows that

$$(7.1) \quad p(r) \geq \alpha(c_\lambda(r)) \quad (0 \leq r \leq R_\lambda).$$

This further implies that if $p(r_0) = \alpha(c_\lambda(r_0))$ for some $r_0 \in (0, R_\lambda)$, then $p'(r_0) = \alpha'(c_\lambda(r_0))c'_\lambda(r_0) > 0$. It is now easy to show that for any $r \in (0, R_\lambda)$, $p'(r)$ cannot

be zero. Indeed, if $p'(r) = 0$ at some point r in $(0, R_\lambda)$, then $p(r) = \alpha(c_\lambda(r))$, by equation (2.21), so that $p'(r) > 0$ by the preceding observation, which is a contradiction. Hence $p'(r) > 0$ for all $0 < r < R_\lambda$.

Finally, we prove assertion (5). Since $u(r) < 0$ for $0 < r < R_\lambda$ and $u(R_\lambda) = 0$, we see that $u'(R_\lambda) \geq 0$. Assume that $u'(R_\lambda) = 0$. Then $u'(R_\lambda) + (2/R_\lambda)u(R_\lambda) = 0$, so that, by assertion (3),

$$u'(r) + \frac{2}{r}u(r) < 0$$

for all $0 < r < R_\lambda$. Multiplying both sides by r^2 and integrating over $[r, R_\lambda]$ yields $r^2 u(r) > 0$ for all $r \in (0, R_\lambda)$, which is a contradiction to (4). Hence $u'(R_\lambda) > 0$. Similarly we can prove that $u'(0) < 0$. \square

Remark 7.1. (1) The conditions $p \in C^1[0, R]$, $u \in C^1[0, R]$ imposed in Theorem 7.1 can be replaced with the weaker conditions $p \in C[0, R] \cap C^1(0, R)$, $u \in C[0, R] \cap C^1(0, R)$. Actually, these conditions can be weakened even further. See Corollary 8.2 in the next section.

(2) From the above proof we see that the inequality (7.1) is strict for $0 < r < R_\lambda$.

If $\lambda \in [\lambda_\infty, 1)$, then the solution of the initial value problem (2.19)–(2.21) satisfies $u'(0) = \beta(\lambda) \geq 0$, and thus, by Theorem 7.1, it cannot be a solution of the free boundary value problem (2.19)–(2.22). Therefore, in the sequel we shall only consider solutions $(p_{\lambda\psi}, u_{\lambda\psi})$ for $\lambda \in (0, \lambda_\infty)$.

We write for brevity

$$p_{\lambda\psi}(r) = p(r), \quad u_{\lambda\psi}(r) = u(r) \quad (0 < r \leq \delta).$$

Since $p(0) = \alpha(\lambda) \in (0, 1)$, $u(0) = 0$ and $u'(0) = \beta(\lambda) < 0$, we can extend (in a unique way) the solution to a maximal interval $0 \leq r \leq R$ such that

$$0 < p(r) < 1, \quad u(r) < 0$$

for all $0 < r < R$, and either $R = R_\lambda$, or $R < R_\lambda$ and one of the following three cases occurs:

- (i) $u(R) = 0$; (ii) $p(R) = 1$; (iii) $p(R) = 0$.

We shall call (p, u) together with the maximal interval $[0, R]$ a *semi-entire solution* of the initial value problem (2.19)–(2.21). The following result extends Theorem 7.1 to semi-entire solutions of the initial value problem.

Theorem 7.2. *Let $\lambda \in (0, \lambda_\infty)$, and let (p, u, R) be a semi-entire solution of the problem (2.19)–(2.21). Then one of the following four situations occurs:*

- (a) $R \leq R_\lambda$ and

$$\begin{aligned} \alpha(c_\lambda(r)) < p(r) < \alpha(c_\lambda(R)), \quad p'(r) > 0 \text{ for } 0 < r < R, \\ u(r) < 0 \text{ for } 0 < r < R, \\ u(R) &= 0; \end{aligned}$$

furthermore, $p(R) = \alpha(c_\lambda(R)) < 1$ if $R < R_\lambda$, and $p(R) = 1$ if $R = R_\lambda$.

- (b) $R \leq R_\lambda$ and

$$\begin{aligned} \alpha(c_\lambda(r)) < p(r) < 1, \quad p'(r) > 0 \text{ for } 0 < r < R, \\ u(r) < 0 \text{ for } 0 < r \leq R, \\ p(R) &= 1. \end{aligned}$$

(c) $R \leq R_\lambda$ and there exists $r_0 \in (0, R)$ such that

$$\begin{aligned} \alpha(c_\lambda(r)) &< p(r) < \alpha(c_\lambda(r_0)) \text{ for } 0 < r < r_0, \\ 0 < p(r) &< \alpha(c_\lambda(r_0)) \text{ for } r_0 < r < R, \\ p'(r) &\begin{cases} > 0 \text{ for } 0 \leq r < r_0, \\ = 0 \text{ for } r = r_0, \\ < 0 \text{ for } r_0 < r \leq R, \end{cases} \\ u(r) &< 0 \text{ for } 0 < r \leq R; \end{aligned}$$

furthermore, either $R < R_\lambda$, $p(R) = 0$ or $R = R_\lambda$, $0 \leq p(R) < 1$.

(d) $R \leq R_\lambda$ and

$$\begin{aligned} 0 < p(r) &< \alpha(\lambda), \quad p'(r) < 0 \text{ for } 0 < r < R, \\ u(r) &< 0 \text{ for } 0 < r \leq R; \end{aligned}$$

furthermore, either $R < R_\lambda$, $p(R) = 0$ or $R = R_\lambda$, $0 \leq p(R) < 1$.

Proof. If $u(R) = 0$, then similar arguments as in the proof of Theorem 7.1 show that Case (a) holds. It remains to consider the case $u(R) < 0$. If $p'(r) \geq 0$ for all $0 < r < R$, then Case (b) follows as in Case (a). Suppose next that $p'(r_0) < 0$ for some $r_0 \in (0, R)$. As before, we denote by (r_1, r_2) the maximal interval in $[0, R]$ containing r_0 such that $p'(r) < 0$ for all $r_1 < r < r_2$. We claim that $r_2 = R$. Indeed, if $r_2 < R$, then $p'(r_2) = 0$, so that by (2.21),

$$(7.2) \quad p(r_2) = \alpha(c_\lambda(r_2)).$$

Since we also have $u(r_1)p'(r_1) = 0$, no matter whether $r_1 > 0$ or $r_1 = 0$, it follows that

$$(7.3) \quad p(r_1) = \alpha(c_\lambda(r_1)).$$

However, (7.2) and (7.3) contradict the fact that $p'(r) < 0$ for $r_1 < r < r_2$. Hence $r_2 = R$. But, then if $r_1 = 0$, then Case (d) holds, whereas if $r_1 > 0$, then, by a similar argument as before, $p'(r) > 0$ for all $0 < r < r_1$, and Case (c) holds. \square

Corollary 7.3. *In cases (a) and (b) of Theorem 7.2, we have*

$$\frac{d}{dr} \left(u'(r) + \frac{2}{r} u(r) \right) > 0 \text{ for } 0 < r < R.$$

Furthermore, in case (a) we also have that $u'(R) > 0$.

The proof is immediate.

Remark 7.2. From (2.19) we notice that, in case (a),

$$u'(R) = -K_D(c_\lambda(R)) + K_M(c_\lambda(R))\alpha(c_\lambda(R)),$$

so that $\alpha(c_\lambda(R)) > K_D(c_\lambda(R))/K_M(c_\lambda(R))$. This gives a lower estimate for R .

A semi-entire solution (p, u, R) with $R < R_\lambda$ can be uniquely extended to a larger interval $[0, R + \delta]$ for some $\delta > 0$. To see this we consider the system of equations (2.19), (2.21) on the interval $[R, R + \delta]$, with initial values $p(R)$ and $u(R)$. Suppose first that (p, u, R) is as in case (a) of Theorem 7.2. Then $p(R) = \alpha(c_0)$ ($c_0 \equiv c_\lambda(R)$), $u(R) = 0$, and $u'(R) > 0$ (by Corollary 7.3). Furthermore, as will be shown in §8 (see (8.10)), $p'(R)$ exists and, in fact,

$$(7.4) \quad p'(R) = \frac{\{K'_P(c_0) + (K'_M(c_0) - K'_N(c_0))\alpha(c_0) - K'_M(c_0)\alpha^2(c_0)\}c'_\lambda(R)}{u'(R) + 2K_M(c_0)\alpha(c_0) - (K_M(c_0) - K_N(c_0))}.$$

Hence we can make the transformation

$$p(r) = \alpha(c_0) + (r - R)(p'(R) + P(r)), \quad u(r) = (r - R)(u'(R) + U(r))$$

with P, U satisfying $P(R) = 0$ and $U(R) = 0$. We then get a system of differential equations which is, by the fact $u'(R) > 0$, similar to that considered in Lemma 5.1 (1) and Theorem 5.3 (1). By similar arguments as before we can then extend the solution uniquely with $u(r) > 0$ to $R < r \leq R + \delta$. Notice that, by (7.4) and the last two equalities of (4.10), we clearly have $p'(R) > 0$, so that if δ is small enough, then $p'(r) > 0$ for $R < r \leq R + \delta$, which further implies that $\alpha(c_0) < p(r) < 1$ for $R < r \leq R + \delta$.

Suppose next that (p, u, R) is as in one of the cases (b)–(d). Then, since $u(R) \neq 0$, the system of equations (2.19) and (2.21) is a regular system at the point $r = R$. We can therefore apply the classical theory of ODEs to uniquely extend the solution to a larger interval $[0, R + \delta]$, although the condition $0 \leq p(r) \leq 1$ may no longer be satisfied for $R < r \leq R + \delta$.

The following theorem asserts that the semi-entire solution can be further extended to the entire interval $0 \leq r \leq R_\lambda$, unless it blows up at some point $\bar{R} \in (R + \delta, R_\lambda]$.

Theorem 7.4. *Let $\lambda \in (0, \lambda_\infty)$ and let (p, u, R) be a semi-entire solution of the system (2.19)–(2.21). Assume that $R < R_\lambda$. Then we have the following conclusions:*

(1) *If (p, u, R) is as in case (a) of Theorem 7.2, then $(p(r), u(r))$ can be extended to the entire interval $[0, R_\lambda]$, satisfying*

$$(7.5) \quad \alpha(c_\lambda(R)) < p(r) < \alpha(c_\lambda(r)), \quad p'(r) > 0 \quad \text{and} \quad u(r) > 0$$

for $R < r \leq R_\lambda$.

(2) *If (p, u, R) is as in case (b), then $(p(r), u(r))$ can be extended either to the entire interval $[0, R_\lambda]$, satisfying*

$$(7.6) \quad p(r) > 1, \quad p'(r) > 0 \quad \text{and} \quad u(r) < 0$$

for $R < r \leq R_\lambda$, or to a half-open interval $[0, \bar{R})$ ($R < \bar{R} \leq R_\lambda$) such that (7.6) holds for $R < r < \bar{R}$, and

$$(7.7) \quad \lim_{r \rightarrow \bar{R}} p(r) = +\infty, \quad \lim_{r \rightarrow \bar{R}} u(r) = 0.$$

(3) *If (p, u, R) is as in cases (c) and (d), then $(p(r), u(r))$ can be extended to the entire interval $[0, R_\lambda]$, satisfying*

$$(7.8) \quad \min_{R \leq r \leq R_\lambda} \alpha_-(c_\lambda(r)) \leq p(r) < 0 \quad \text{and} \quad u(r) < 0$$

for $R < r \leq R_\lambda$, where α_- is the negative root of the equation (3.1), i.e.,

$$(7.9) \quad \alpha_-(\lambda) = \frac{1}{2K_M(\lambda)} \left(K_M(\lambda) - K_N(\lambda) - \sqrt{(K_M(\lambda) - K_N(\lambda))^2 + 4K_M(\lambda)K_P(\lambda)} \right).$$

Proof. Consider Case (1) first. Let (R, r_0) be the largest open interval contained in (R, R_λ) such that the solution exists and $p'(r) > 0$ and $u(r) > 0$ for all $r \in (R, r_0)$. We claim that $r_0 = R_\lambda$ and $(p(r), u(r))$ is well-defined at $r = R_\lambda$. Indeed, by (2.21) and (2.19), the conditions $p'(r) > 0$ and $u(r) > 0$ imply that

$$\alpha(c_\lambda(R) < p(r) < \alpha(c_\lambda(r)), \quad \left| u'(r) + \frac{2}{r}u(r) \right| \leq \text{const.} < \infty$$

for $R < r < r_0$, so that $p(r_0) = \lim_{r \rightarrow r_0-0} p(r)$ and $u(r_0) = \lim_{r \rightarrow r_0-0} u(r)$ exist. Furthermore, since $p'(r) > 0$, by (2.19) we see that $u'(r) + (2/r)u(r)$ is monotone increasing in (R, r_0) , which implies that $(r^2 u(r))' > r^2(u'(R) + (2/R)u(R)) > 0$ and, therefore, $u(r_0) > 0$. It follows that the system of equations (2.19), (2.21) is regular at r_0 , so that $p'(r_0)$ exists. We claim that $p'(r_0) > 0$. Indeed, if $p'(r_0) = 0$, then $p(r_0) = \alpha(c_\lambda(r_0))$. But then, since $p(r) < \alpha(c_\lambda(r))$ for $R < r < r_0$, we infer that $p'(r_0) \geq \alpha'(c_\lambda(r_0))c'_\lambda(r_0) > 0$, which is a contradiction. Hence $p'(r_0) > 0$. It follows that if $r_0 < R_\lambda$, then we can extend (p, u) to a larger interval on which the conditions $p'(r) > 0$ and $u(r) > 0$ still hold, which is contrary to the maximality of r_0 . Hence $r_0 = R_\lambda$, and Case (1) follows.

Consider Case (2). Since $p(R) = 1$ and $u(R) < 0$, it is easy to see (by (2.21)) that $p'(R) > 0$, so that (7.6) holds for r larger than but near R . Let (R, r_0) be the largest open interval such that the solution exists and $p'(r) > 0$, $u(r) < 0$ for all $R < r < r_0$. Then either $p(r_0) = \lim_{r \rightarrow r_0-0} p(r)$ exists and is greater than 1, or $\lim_{r \rightarrow r_0-0} p(r) = +\infty$. In the first case it follows from (2.19) that $u(r_0) = \lim_{r \rightarrow r_0-0} u(r)$ also exists. Since $p'(r) > 0$ for $R < r < r_0$, we have

$$p(r_0) > p(R) = 1 \geq \alpha(c_\lambda(r_0)),$$

so that the right-hand side of (2.21) is $\neq 0$; hence $u(r_0) \neq 0$ and, therefore, $u(r_0) < 0$ and $p'(r_0) > 0$. But then $r_0 = R_\lambda$, for otherwise we can extend (p, u) to a larger interval such that (7.6) still holds, which contradicts the maximality of (r, r_0) . Consider next the second case, where $p(r_0-0) = +\infty$. We claim that

$$(7.10) \quad \lim_{r \rightarrow r_0} u(r) = 0.$$

Indeed, since $u(r) < 0$ for $r < r_0$, by (2.19) we have

$$u'(r) > -K_D(c_\lambda(r)) + K_M(c_\lambda(r))p(r) \rightarrow +\infty$$

as $r \rightarrow r_0$, so that $u(r)$ is monotone increasing for r near r_0 . It follows that $u(r)$ has a limit as r approaches r_0 . We claim that $\kappa \equiv \lim_{r \rightarrow r_0} u(r)$ is equal to zero. Indeed, suppose $\kappa < 0$ and let

$$y(r) = \begin{cases} \frac{1}{p(r)} & \text{for } 0 < r < r_0, \\ 0 & \text{for } r = r_0. \end{cases}$$

Clearly, $y \in C(0, r_0] \cap C^1(0, r_0)$, and by (2.21),

$$\begin{aligned} \lim_{r \rightarrow r_0} y'(r) &= - \lim_{r \rightarrow r_0} \frac{K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda(r)))p(r) - K_M(c_\lambda(r))p^2(r)}{u(r)p^2(r)} \\ &= \frac{K_M(c_\lambda(r_0))}{\kappa}. \end{aligned}$$

Hence

$$y(r) = -\frac{K_M(c_\lambda(r_0))}{\kappa}(r_0 - r) + o(r_0 - r) \quad \text{as } r \rightarrow r_0,$$

so that

$$p(r) = -\frac{\kappa}{K_M(c_\lambda(r))} \cdot \frac{1}{r_0 - r} (1 + o(1)) \quad \text{as } r \rightarrow r_0.$$

By (2.19), we then get

$$u'(r) = -\frac{\kappa}{r_0 - r} (1 + o(1)) \quad \text{as } r \rightarrow r_0,$$

so that $u(r) \rightarrow -\infty$ as $r \rightarrow r_0$, which is a contradiction. This proves (7.10). Thus assertion (2) holds with $\bar{R} = r_0$.

Finally, we consider case (3). Since $p(R) = 0$ and $u(R) < 0$, by (2.21) it follows that $p'(R) < 0$, so that $\alpha_-(r) < p(r) < 0$, $p'(r) < 0$ and $u(r) < 0$ for r larger than but near R . Let (R, r_0) be the largest open interval such that the solution exists and

$$p(r) < 0, \quad u(r) < 0$$

for $R < r < r_0$. From (2.21) we infer that

$$p'(r) \begin{cases} < 0 & \text{if } \alpha_-(c_\lambda(r)) < p(r) < 0, \\ > 0 & \text{if } p(r) < \alpha_-(c_\lambda(r)), \end{cases}$$

for all $R < r < r_0$, and this implies that

$$p(r) \geq \min_{R \leq r \leq r_0} \alpha_-(c_\lambda(r)) \geq \min_{R \leq r \leq R_\lambda} \alpha_-(c_\lambda(r))$$

for $R < r < r_0$, so that $p(r)$ is bounded. Using (2.19), we find that $u(r_0) \equiv \lim_{r \rightarrow r_0} u(r)$ exists and is negative and $u(r)$ has a negative upper bound in the interval $R < r < r_0$. Hence, by (2.21), $p'(r)$ is bounded for $R < r < r_0$, which implies that also $p(r_0) \equiv \lim_{r \rightarrow r_0} p(r)$ exists. We claim that $p(r_0) < 0$. Indeed, if $p(r_0) = 0$, then on the one hand, since $p(r) < 0$ for $R < r < r_0$, $p'(r_0) \geq 0$; but on the other hand, it follows from (2.21) and the fact $u(r_0) < 0$ that $p'(r_0) < 0$, which is a contradiction. Recalling the maximality of (R, r_0) , we conclude that $r_0 = R_\lambda$, and (7.8) holds for all $R < r \leq R_\lambda$. \square

Remark 7.3. Every entire solution is analytic at all points except possibly at $r = 0$. Indeed, at a point where u does not vanish, the analyticity of p and u follows immediately from classical results, and where u vanishes, the analyticity follows as in the proof of Theorem 4.1, noting that if $u(r_0) = 0$ and $r_0 > 0$, then $u'(r_0) > 0$ so that in the equations analogous to (4.9) the coefficients of $p^{(k)}(r_0)$ are uniformly positive.

Remark 7.4. We can prove that if (p, u) is as in (7.7), then there exists a constant $\kappa > 0$ such that

$$p(r) = \frac{\kappa}{\sqrt{\bar{R} - r}}(1 + o(1)), \quad u(r) = -2K_M(c_\lambda(\bar{R}))\kappa\sqrt{\bar{R} - r}(1 + o(1))$$

as $r \rightarrow \bar{R}$. Since we shall not use this result, we omit the proof.

Definition 7.1. The solutions obtained in Theorem 7.4 are called *entire solutions*. We distinguish four types of such solutions in accordance with Theorem 7.4:

(i) If $0 < p(r) \leq 1$, $p'(r) > 0$ for $0 < r \leq R_\lambda$, and there exists an $R \in (0, R_\lambda]$ such that $u(r) < 0$ for $0 < r < R$, $u(R) = 0$, and $u(r) > 0$ for $R < r \leq R_\lambda$, then we call the solution of the initial value problem (2.19)–(2.21) a *subsolution* of the free boundary problem (2.19)–(2.22); in particular, if $R < R_\lambda$, then we call it a *strict subsolution*.

(ii) If $0 < p(r) < 1$, $p'(r) > 0$ and $u(r) < 0$ for $0 < r < R_\lambda$ and $p(R_\lambda) = 1$, $u(R_\lambda) < 0$ ($\implies p'(R_\lambda) = 0$), then we call it a *supersolution* of the free boundary problem (2.19)–(2.22).

(iii) If $p'(r) > 0$, $u(r) < 0$ for $0 < r \leq R_\lambda$, and there exists $R \in (0, R_\lambda]$ such that $0 < p(r) < 1$ for $0 < r < R$, $p(R) = 1$, and $p(r) > 1$ for $R < r \leq R_\lambda$, then we call it a *bounded upper solution*, whereas if there exists $\bar{R} \in (0, R_\lambda]$ such that (p, u) is

only defined for $0 < r < \bar{R}$, $p'(r) > 0$, $u(r) < 0$ for $0 < r < \bar{R}$, and (7.7) holds, then we call it an *unbounded upper solution*.

(iv) If the solution is as in Theorem 7.4 (3), or equivalently, $p'(r_0) < 0$ for some $r_0 \in (0, R_\lambda)$, then we call it a *lower solution*.

Note that, by the definition, a supersolution is also a bounded upper solution, and a solution of the free boundary problem (2.19)–(2.22) is a subsolution.

In §§9, 10 we shall prove that, for each $\lambda \in (0, \lambda_\infty)$, there exists a unique $\bar{\psi} = \bar{\psi}(\lambda) \in \mathbf{R}$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is an upper solution for $\psi > \bar{\psi}$ and a lower solution for $\psi < \bar{\psi}$, whereas $(p_\lambda, u_\lambda) \equiv (p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is either a subsolution or a supersolution. Moreover, if $(p_{\lambda_1}, u_{\lambda_1})$ is a subsolution and $(p_{\lambda_2}, u_{\lambda_2})$ is a supersolution, then there exists a λ^* between λ_1 and λ_2 such that $(p_{\lambda^*}, u_{\lambda^*})$ is a solution of the free boundary problem (2.19)–(2.22).

8. WEAK SOLUTIONS OF THE INITIAL VALUE PROBLEM

In this section we reformulate the problem (2.19)–(2.21) as a system of integral equations, and use it to introduce the concept of weak solutions. The integral equation formulation presented here will enable us to work with weak limits of solutions.

Theorem 8.1. *Suppose that $p(r), u(r) \in L^\infty[0, R] \cap C^1(0, R]$ and they satisfy equations (2.19), (2.21) for $0 < r \leq R$, where $0 < R \leq R_\lambda$, and R_λ is as in (2.18). Then $p(r), u(r)$ also satisfy the equations*

$$(8.1) \quad u(r) = \frac{1}{r^2} \int_0^r \left(-K_D(c_\lambda(\rho)) + K_M(c_\lambda(\rho))p(\rho) \right) \rho^2 d\rho,$$

$$(8.2) \quad u(r)p(r) = \frac{1}{r^2} \int_0^r \left(K_P(c_\lambda(\rho)) + (K_B(c_\lambda(\rho)) - K_N(c_\lambda(\rho)))p(\rho) \right) \rho^2 d\rho$$

for $0 < r \leq R$. Conversely, if $p(r), u(r)$ belong to $L^\infty[0, R]$ and satisfy the system of equations (8.1), (8.2) for almost all $0 < r \leq R$, and

$$(8.3) \quad p(r) \geq 0 \quad \text{for } 0 \leq r \leq R,$$

then, after modifying the values of u and p on a subset of measure zero, $p \in C[0, R] \cap C^1(0, R]$, $u \in C^1[0, R]$, and they satisfy (2.19)–(2.21) for all $0 < r \leq R$; furthermore,

(i) $p(0)$ satisfies the equation

$$(8.4) \quad K_P(\lambda) + (K_M(\lambda) - K_N(\lambda))p(0) - K_M(\lambda)p^2(0) = 0;$$

(ii) if $\gamma_1(\lambda) > 0$, then $p(r)$ is continuously differentiable at $r = 0$, and $p'(0) = 0$; if $\gamma_1(\lambda) = 0$, then $p(r)$ is continuously differentiable at $r = 0$, but $p'(0)$ can be any real number; if $\gamma_1(\lambda) < 0$ ($\iff -1 < \sigma(\lambda) < 0$), then the limit

$$\omega = \lim_{r \rightarrow 0} p'(r)r^{-\sigma(\lambda)}$$

exists, and if $\omega = 0$, then $p(r)$ is continuously differentiable at $r = 0$ and $p'(0) = 0$.

Proof. Suppose first that $p(r), u(r)$ satisfy (2.19), (2.21). Multiplying (2.19) with r^2 and integrating over $[0, r]$ (for arbitrary $0 < r \leq R$) immediately yields (8.1). Next, multiplying (2.21) with r^2 , integrating over $[0, r]$, and using integration by parts and equation (2.19), we obtain (8.2).

Now suppose that $p, u \in L^\infty[0, R]$ and they satisfy (8.1)–(8.3) for almost all $0 < r \leq R$. We modify the values of $u(r)$ on a subset of measure zero by replacing it with the right-hand side of (8.1). Then $u(r)$ is continuous for all $0 \leq r \leq R$, and $u(0) = 0$. Similarly, after modifying the values of $p(r)$ on a subset of measure zero of the open set

$$(8.5) \quad O = \{r \in (0, R) : u(r) \neq 0\},$$

we may assume that $p(r)$ satisfies (8.2) and is continuous at all $r \in O$. From (8.1), (8.2) we further conclude that $u(r), p(r)$ are continuously differentiable in O . Let

$$(8.6) \quad \Sigma = \{r \in [0, R] : u(r) = 0\}.$$

We claim that

$$(8.7) \quad \text{meas}(\Sigma) = 0.$$

Indeed, if $\text{meas}(\Sigma) > 0$, then $\text{meas}(\Sigma^*) = \text{meas}(\Sigma) > 0$, where Σ^* is the set of Lebesgue points of Σ . Since at each point of Σ we have

$$\begin{aligned} \int_0^r \left(-K_D(c_\lambda(\rho)) + K_M(c_\lambda(\rho))p(\rho) \right) \rho^2 d\rho &= 0 \quad (\text{by (8.1)}), \\ \int_0^r \left(K_P(c_\lambda(\rho)) + (K_B(c_\lambda(\rho)) - K_N(c_\lambda(\rho)))p(\rho) \right) \rho^2 d\rho &= 0 \quad (\text{by (8.2)}), \end{aligned}$$

it follows that, for almost all $r \in \Sigma^*$,

$$\begin{aligned} -K_D(c_\lambda(r)) + K_M(c_\lambda(r))p(r) &= 0, \\ K_P(c_\lambda(r)) + (K_B(c_\lambda(r)) - K_N(c_\lambda(r)))p(r) &= 0. \end{aligned}$$

Eliminating $p(r)$, we get

$$\varphi(r) \equiv K_M(c_\lambda(r))K_P(c_\lambda(r)) + (K_B(c_\lambda(r)) - K_N(c_\lambda(r)))K_D(c_\lambda(r)) = 0$$

a.e. in Σ^* . On the other hand, if we denote by c^* the unique positive number such that $K_B(c^*) = K_Q(c^*)$, then by writing $\varphi = K_B K_P + (K_B - K_Q)K_D$, we easily verify that $\varphi'(r) > 0$ if $c_\lambda(r) \leq c^*$ and $\varphi(r) > 0$ if $c_\lambda(r) \geq c^*$, so that $\varphi(r)$ has at most one zero, contradicting the fact that $\varphi = 0$ on a set of positive measure. This completes the proof of (8.7). Since $\text{meas}(\Sigma) = 0$, we may redefine $p(r)$ for $r \in \Sigma$ by

$$(8.8) \quad p(r) = \alpha(c_\lambda(r)) \quad (r \in \Sigma);$$

this does not change the fact that both (8.1) and (8.2) are satisfied for all $0 < r \leq R$. With the definition (8.8) at hand, we shall be able to prove that $p(r)$ is continuous for $r \in \Sigma$.

Differentiating (8.1) and (8.2), one easily finds that $u(r), p(r)$ satisfy equations (2.19) and (2.21) in O . Clearly, O consists of a countable number of disjoint open intervals. Let (r_1, r_2) be any one of these open intervals (so that $u(r_1) = 0$, and $u(r_2) = 0$ if $r_2 < R$). Arguing similarly as in the proof of Theorem 7.2, we infer that $p'(r)$ changes sign at most once in (r_1, r_2) . It follows that $p(r)$ has limits as $r \rightarrow r_1 + 0$ and $r \rightarrow r_2 - 0$. By (8.1), (8.2) we have

$$p(r) = \frac{\int_0^r \left(K_P(c_\lambda(\rho)) + (K_B(c_\lambda(\rho)) - K_N(c_\lambda(\rho)))p(\rho) \right) \rho^2 d\rho}{\int_0^r \left(-K_D(c_\lambda(\rho)) + K_M(c_\lambda(\rho))p(\rho) \right) \rho^2 d\rho}$$

for $r_1 < r < r_2$. Using l'Hospital's rule, we deduce that

$$(8.9) \quad K_P(c_\lambda(r_1)) + (K_M(c_\lambda(r_1)) - K_N(c_\lambda(r_1)))p(r_1 + 0) - K_M(c_\lambda(r_1))p^2(r_1 + 0) = 0.$$

By Lemma 3.1 and (8.8), it follows that

$$p(r_1 + 0) = \alpha(c_\lambda(r_1)) = p(r_1).$$

As for $p(r_2 - 0)$, if either $r_2 < R$ ($\implies u(r_2) = 0$) or $r_2 = R$ and $u(R) = 0$, then a similar argument as above shows that $p(r_2 - 0) = p(r_2) = \alpha(c_\lambda(r_2))$. If instead $r_2 = R$ and $u(R) \neq 0$, then clearly $p(R - 0) = p(R)$. Hence, p is right continuous at r_1 and left continuous at r_2 . This further implies that $u'(r_1 + 0)$ and $u'(r_2 - 0)$ exist. Using arguments similar to those used in the proof of Theorem 7.1, we can show, if either $r_2 < R$, or $r_2 = R$ and $u(R) = 0$, that $p'(r) \geq 0$ and $u(r) < 0$ for $r_1 < r < r_2$, and $u'(r_1 + 0) < 0$, $u'(r_2 - 0) > 0$. From the above discussion it follows, in particular, that if r_0 is an isolated point in Σ , then p is continuous at r_0 .

Consider next the case where r_0 is an accumulation point of Σ . Since u cannot be identically zero at any intervals, O is dense in $[0, R]$ so that r_0 is an accumulation point of O . Let $\{r_n\}_{n=1}^\infty$ be a sequence of points in $[0, R]$ converging to r_0 . By splitting it into two distinct subsequences if necessary, we may assume that $\{r_n\}_{n=1}^\infty$ is either monotone increasing or monotone decreasing. Consider the case where $\{r_n\}_{n=1}^\infty$ is monotone increasing; the monotone decreasing case can be dealt with in a similar way. By further splitting the sequence into two subsequences if necessary, we may assume that $\{r_n\}_{n=1}^\infty$ satisfies one of the following two conditions:

- (a) $r_n \in \Sigma$, $n = 1, 2, \dots$;
- (b) $r_n \in O$, $n = 1, 2, \dots$.

In Case (a) we have $p(r_n) = \alpha(c_\lambda(r_n))$, so that

$$\lim_{n \rightarrow \infty} p(r_n) = \lim_{n \rightarrow \infty} \alpha(c_\lambda(r_n)) = \alpha(c_\lambda(r_0)) = p(r_0).$$

In Case (b) two situations may occur:

- (b1) There is $\delta > 0$ such that $(r_0 - \delta, r_0) \subset O$;
- (b2) r_0 is the limit of an increasing sequence of points in Σ .

In the first situation we know by the preceding analysis that $\lim_{r \rightarrow r_0 - 0} p(r) = p(r_0)$, so that

$$\lim_{n \rightarrow \infty} p(r_n) = p(r_0).$$

In the second situation we denote by (r_{1n}, r_{2n}) the open interval contained in O such that $r_{1n} < r_n < r_{2n}$ and $u(r_{1n}) = u(r_{2n}) = 0$, $n = 1, 2, \dots$. Since p is monotone increasing in (r_{1n}, r_{2n}) ,

$$\alpha(c_\lambda(r_{1n})) = p(r_{1n} + 0) < p(r_n) < p(r_{2n} - 0) = \alpha(c_\lambda(r_{2n})), \quad n = 1, 2, \dots,$$

so that

$$\lim_{n \rightarrow \infty} p(r_n) = \lim_{n \rightarrow \infty} \alpha(c_\lambda(r_{1n})) = \lim_{n \rightarrow \infty} \alpha(c_\lambda(r_{2n})) = \alpha(c_\lambda(r_0)) = p(r_0).$$

Hence p is continuous at r_0 .

Having proved that p is continuous at any point of Σ , we conclude that $p \in C[0, R]$ and, consequently, $u \in C^1[0, R]$. (The assertion that $\lim_{r \rightarrow 0} u'(r)$ exists can be easily verified using (8.1) and (2.19).) Furthermore, since p is monotone increasing in every open interval (r_1, r_2) of O such that either $r_2 < R$ or $r_2 = R$

and $u(R) = 0$, and since p is also monotone increasing in Σ , we conclude that if Σ contains more than one point, p is monotone increasing in the interval $[0, r^*]$, where

$$r^* = \sup\{r : r \in \Sigma\};$$

if $r^* < R$, then either p is monotone increasing in $[r^*, R]$ or it changes monotonicity once, say, at the point $r^{**} \in (r^*, R)$, and p is increasing in $[0, r^{**}]$ and decreasing in $[r^{**}, R]$. It follows that $u'(r) + (2/r)u(r)$ is monotone increasing for either $0 < r \leq R$ or $0 < r \leq r^{**}$ (if $r^{**} < R$). This implies, by similar arguments as in the proofs of Theorems 7.1 and 7.2, that either $\Sigma = \{0\}$ or $\Sigma = \{0, r_0\}$ for some $0 < r_0 \leq R$.

We next prove that if $\Sigma = \{0, r_0\}$ ($0 < r_0 \leq R$), then

$$\begin{aligned} (8.10) \quad \lim_{r \rightarrow r_0 - 0} p'(r) &= \frac{\{K'_P(c_0) + (K'_M(c_0) - K'_N(c_0))\alpha(c_0) - K'_M(c_0)\alpha^2(c_0)\}c'_\lambda(r_0)}{u'(r_0) + 2K_M(c_0)\alpha(c_0) - (K_M(c_0) - K_N(c_0))} \\ &\equiv I(r_0), \end{aligned}$$

where $c_0 = c_\lambda(r_0)$. (Note that $I(r_0)$ is the derivative of p at r_0 formally computed by differentiating equation (2.21) at $r = r_0$ and setting $u(r_0)p''(r_0) = 0$.) To this end we differentiate (2.21) for $0 < r < r_0$ and divide it by $u(r)$ to get

$$(8.11) \quad p''(r) + \frac{a(r)}{u(r)}p'(r) = \frac{b(r)}{u(r)},$$

where

$$\begin{aligned} a(r) &= u'(r) + 2K_M(c_\lambda(r))p(r) - (K_M(c_\lambda(r)) - K_N(c_\lambda(r))), \\ b(r) &= \{K'_P(c_\lambda(r)) + (K'_M(c_\lambda(r)) - K'_N(c_\lambda(r)))p(r) - K'_M(c_\lambda(r))p^2(r)\}c'_\lambda(r). \end{aligned}$$

Let \bar{r} be an arbitrary point in $(0, r_0)$ and set

$$A(r) = \int_{\bar{r}}^r \frac{a(\rho)}{u(\rho)} d\rho, \quad 0 < r < r_0.$$

Since $u(r) = u'(r_0)(r - r_0) + o(r - r_0)$ as $r \rightarrow r_0$ and $u'(r_0) > 0$, $a(r_0) > 0$ (by (3.2)), it follows that

$$(8.12) \quad A(r) = \frac{a(r_0)}{u'(r_0)} \log |r - r_0| \cdot (1 + o(1)) \quad \text{as } r \rightarrow r_0 - 0.$$

We multiply (8.11) with $e^{A(r)}$ and integrate over $[\bar{r}, r]$ to get

$$(8.13) \quad p'(r)e^{A(r)} = p'(\bar{r})e^{A(\bar{r})} + \int_{\bar{r}}^r \frac{b(\rho)}{u(\rho)} e^{A(\rho)} d\rho.$$

By (8.12) we have

$$\frac{b(r)}{u(r)} e^{A(r)} = -\frac{b(r_0)}{u'(r_0)} |r - r_0|^{\nu(1+o(1))} (1 + o(1)) \quad \text{as } r \rightarrow r_0 - 0,$$

where

$$\nu = \frac{a(r_0)}{u'(r_0)} - 1 = \frac{2K_M(c_0)\alpha(c_0) - (K_M(c_0) - K_N(c_0))}{u'(r_0)} > 0 \quad (\text{by (3.2)}).$$

It follows that $(b(r)/u(r))e^{A(r)}$ is integrable on $[\bar{r}, r_0)$. Hence, by (8.13),

$$\kappa \equiv \lim_{r \rightarrow r_0 - 0} p'(r)e^{A(r)}$$

exists, and

$$(8.14) \quad \kappa = p'(\bar{r})e^{A(\bar{r})} + \int_{\bar{r}}^{r_0} \frac{b(\rho)}{u(\rho)} e^{A(\rho)} d\rho.$$

We claim that $\kappa = 0$. To see this we assume the converse: $\kappa \neq 0$. Since $p'(r) \geq 0$ for $0 < r < r_0$, we have $\kappa > 0$. From the definition of κ and (8.12) it then follows that

$$(8.15) \quad \lim_{r \rightarrow r_0-0} p'(r) = +\infty.$$

We now write, for $0 < r < r_0$,

$$p'(r) = \frac{K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda(r)))p(r) - K_M(c_\lambda(r))p^2(r)}{u(r)},$$

and by l'Hospital's rule,

$$\begin{aligned} & \lim_{r \rightarrow r_0-0} p'(r) \\ &= \lim_{r \rightarrow r_0-0} \frac{(K'_P + (K'_M - K'_N)p - K'_M p^2)c'_\lambda(r) - (2K_M p - (K_M - K_N))p'(r)}{u'(r)}. \end{aligned}$$

However, by (8.15), the left-hand side is $+\infty$ while the right-hand side is $-\infty$, which is a contradiction.

Having proved that $\kappa = 0$, we infer from (8.13) and (8.14) that

$$(8.16) \quad p'(r) = - \int_r^{r_0} \frac{b(\rho)}{u(\rho)} e^{A(\rho)} d\rho / e^{A(r)},$$

and, invoking l'Hospital's rule again, (8.10) easily follows.

Similarly we can prove that $\lim_{r \rightarrow r_0+0} p'(r) = I(r_0)$. Hence $p(r)$ is continuously differentiable at r_0 .

Finally, assertion (i) follows immediately from (8.9) by taking $r_1 = 0$, and assertion (ii) follows from (8.13) and

$$\frac{a(0)}{u'(0)} = \frac{\gamma_1(\lambda)}{\beta(\lambda)} = -\sigma(\lambda),$$

by using previous arguments. This completes the proof of Theorem 8.1. \square

Corollary 8.2. *Let (p, u, R_λ) be a solution of the free boundary problem (2.19)–(2.22) with $p, u \in C[0, R_\lambda] \cap C^1(0, R_\lambda)$. Then assertions (1)–(5) of Theorem 7.1 hold.*

Proof. Indeed, in the proof of Theorem 7.1, the condition that $p(r)$ is differentiable at $r = 0$ and $r = R_\lambda$ is used only to ensure that at these points $u(r)p'(r) = 0$ so that

$$(8.17) \quad p(0) = \alpha(\lambda) \quad \text{and} \quad p(R_\lambda) = \alpha(1).$$

By Theorem 8.1 we see that (8.17) holds also if $p \in C[0, R_\lambda] \cap C^1(0, R_\lambda)$, $u \in C[0, R_\lambda] \cap C^1(0, R_\lambda)$ and (p, u) satisfy (2.19)–(2.22). Hence the desired assertion follows. \square

Definition 8.1. A pair of functions p, u in $L^\infty[0, R]$ satisfying (8.1), (8.2) a.e. in $[0, R]$ is called a *weak solution* of (2.19)–(2.21).

Theorem 8.1 asserts that if (p, u) is a weak solution and $p \geq 0$, then it is also a classical solution, that is, $p \in C[0, R] \cap C^1(0, R]$, $u \in C^1[0, R]$, and (2.19)–(2.21)

hold. However, if we remove the condition $p \geq 0$, then the result may be different, as briefly discussed in the next paragraph.

The condition (8.3) was used to prove that $p(r_1 - 0) = \alpha(c_\lambda(r_1))$ and similarly that $p(r_1 + 0) = \alpha(c_\lambda(r_1))$, so that $p(r)$ is continuous at any boundary point r_1 of O . If $p(r)$ changes sign, then it may occur that

$$p(r_1 - 0) = \alpha(c_\lambda(r_1)) \quad \text{but} \quad p(r_1 + 0) = \alpha_-(c_\lambda(r_1))$$

(or vice versa), where $\alpha_-(\lambda)$ is the negative root of (3.1) (see (7.9)). Thus a weak solution need not be a classical solution.

The following result supplements Theorem 8.1 in case (p, u) is a weak solution and p may change sign.

Theorem 8.3. *Let (p, u) and (\bar{p}, \bar{u}) be respectively a weak solution and a classical solution of the problem (2.19)–(2.21) with the same λ , $\bar{p} \geq 0$. Suppose that*

$$(8.18) \quad p(r) = \bar{p}(r), \quad u(r) = \bar{u}(r)$$

for $0 < r < \delta$, for some $\delta > 0$. Then (8.18) holds for all $0 < r < R$, where R is equal to:

- (i) the positive zero of \bar{u} if (\bar{p}, \bar{u}) is a subsolution,
- (ii) the blow-up point of \bar{p} if (\bar{p}, \bar{u}) is an unbounded upper solution,
- (iii) R_λ if (\bar{p}, \bar{u}) is a bounded upper solution, or
- (iv) the first positive zero of $\frac{d\bar{p}(r)}{dr}$ if (\bar{p}, \bar{u}) is a lower solution.

Proof. Let $(0, r_0)$ ($r_0 > 0$) be the largest open interval such that (8.18) holds for all $r \in (0, r_0)$. We claim that $r_0 \geq R$. Indeed, if $r_0 < R$, then, since (8.18) holds for all $r \in (0, r_0)$ and since $u(r)$ is a continuous function (by (8.1)) and $\bar{u}(r) < 0$ for $0 < r < R$, it follows that $u(r) < 0$ for $0 < r < r_0 + \varepsilon$ (for some $\varepsilon > 0$; $r_0 + \varepsilon < R$). Thus, by the proof of Theorem 8.1, $p(r)$ is continuous and $(p(r), u(r))$ is a classical solution for $0 < r < r_0 + \varepsilon$. By uniqueness of classical solutions we deduce that $p(r) = \bar{p}(r)$ and $u(r) = \bar{u}(r)$ for $0 < r < r_0 + \varepsilon$, which contradicts the maximality of r_0 . \square

9. EXISTENCE OF SUBSOLUTIONS

In this section we prove that for any λ less than, but near, λ_∞ , there exists a unique $\bar{\psi} = \bar{\psi}(\lambda)$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution. In the next section we shall prove that for any λ near 0, there exists a unique $\bar{\psi} = \bar{\psi}(\lambda)$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a supersolution. These facts will be used to prove the existence of a solution to the free boundary problem.

The following comparison lemma will play a crucial role:

Lemma 9.1. *Let (p_1, u_1) and (p_2, u_2) be two different solutions of (2.19)–(2.21) with the same $\lambda \in (0, \lambda_\infty)$, defined on the same interval $[0, R)$ ($0 < R \leq R_\lambda$). Suppose that*

$$(9.1) \quad p_1(r) > p_2(r) \quad \text{and} \quad u_1(r) > u_2(r)$$

in some small interval $0 < r < \delta$. Suppose further that

$$(9.2) \quad u_1(r) < 0, \quad u_2(r) < 0 \quad \text{and} \quad p'_2(r) > 0$$

for $0 < r < R$. Then (9.1) holds for all $0 < r < R$.

Proof. Assume that the assertion is not true. Then there exists an $r_0 \in (0, R)$ such that (9.1) holds for $0 < r < r_0$ and either (i) $p_1(r_0) > p_2(r_0)$, $u_1(r_0) = u_2(r_0)$, or (ii) $p_1(r_0) = p_2(r_0)$, $u_1(r_0) > u_2(r_0)$. (Note that the case $p_1(r_0) = p_2(r_0)$, $u_1(r_0) = u_2(r_0)$ cannot take place, by uniqueness of solutions.) In the first case we have $u'_1(r_0) > u'_2(r_0)$, by (2.19). On the other hand, from $u_1(r) > u_2(r)$ for $0 < r < r_0$ and $u_1(r_0) = u_2(r_0)$, it follows that $u'_1(r_0) \leq u'_2(r_0)$, which is a contradiction. In the second case we have $u_1(r_0)p'_1(r_0) = u_2(r_0)p'_2(r_0)$, by (2.21). Since $|u_2(r_0)| > |u_1(r_0)|$, we get

$$\frac{p'_1(r_0)}{p'_2(r_0)} = \frac{u_2(r_0)}{u_1(r_0)} > 1,$$

so that $p'_1(r_0) > p'_2(r_0)$. On the other hand, from $p_1(r) > p_2(r)$ for $0 < r < r_0$ and $p_1(r_0) = p_2(r_0)$, it follows that $p'_1(r_0) \leq p'_2(r_0)$, which is again a contradiction. Hence the desired assertion follows. \square

Lemma 9.2. *For any $\lambda \in (0, \lambda_\infty)$ the following hold:*

- (1) *If for some $\bar{\psi} \in \mathbf{R}$, $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution, then for all $\psi > \bar{\psi}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is an unbounded upper solution, and for all $\psi < \bar{\psi}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is a lower solution.*
- (2) *If for some $\bar{\psi} \in \mathbf{R}$, $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is an upper solution, then for all $\psi > \bar{\psi}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is also an upper solution; in particular, if for some $\bar{\psi} \in \mathbf{R}$, $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is an unbounded upper solution, then for all $\psi > \bar{\psi}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is also an unbounded upper solution.*
- (3) *If for some $\bar{\psi} \in \mathbf{R}$, $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a lower solution, then for all $\psi < \bar{\psi}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is also a lower solution.*

Proof. Assertion (2) follows quickly from Lemma 9.1. Indeed, from the expressions (5.43)–(5.46) and (5.56) near $r = 0$ and the fact that $\omega = \omega(\lambda, \psi)$ is monotone increasing in ψ it follows that, if $\psi > \bar{\psi}$, then

$$(9.3) \quad p_{\lambda\psi}(r) > p_{\lambda\bar{\psi}}(r), \quad u_{\lambda\psi}(r) > u_{\lambda\bar{\psi}}(r)$$

for r near 0. If $(p_{\lambda\bar{\psi}}(r), u_{\lambda\bar{\psi}}(r))$ is an upper solution, then $\frac{d}{dr}p_{\lambda\bar{\psi}}(r) > 0$, so that by Lemma 9.1, the above inequalities hold also for all r such that both $(p_{\lambda\psi}(r), u_{\lambda\psi}(r))$ and $(p_{\lambda\bar{\psi}}(r), u_{\lambda\bar{\psi}}(r))$ are well-defined.

To prove assertion (1), consider first the case $\psi > \bar{\psi}$. As before, (9.3) holds for all r such that both $(p_{\lambda\psi}(r), u_{\lambda\psi}(r))$ and $(p_{\lambda\bar{\psi}}(r), u_{\lambda\bar{\psi}}(r))$ are well-defined. It follows that $(p_{\lambda\psi}, u_{\lambda\psi})$ is either a subsolution or an upper solution. We claim that $(p_{\lambda\psi}, u_{\lambda\psi})$ cannot be a subsolution. Indeed, if $(p_{\lambda\psi}, u_{\lambda\psi})$ is a subsolution, then, by (9.3), the zero $\tilde{R}_{\lambda\psi}$ of $u_{\lambda\psi}(r)$ and the zero $\tilde{R}_{\lambda\bar{\psi}}$ of $u_{\lambda\bar{\psi}}(r)$ must satisfy the inequality $\tilde{R}_{\lambda\psi} < \tilde{R}_{\lambda\bar{\psi}}$. Since $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution, $p_{\lambda\bar{\psi}}(r) > \alpha(c_\lambda(r))$ for $0 < r < \tilde{R}_{\lambda\bar{\psi}}$, so that

$$p_{\lambda\psi}(\tilde{R}_{\lambda\psi}) = \alpha(c_\lambda(\tilde{R}_{\lambda\psi})) < p_{\lambda\bar{\psi}}(\tilde{R}_{\lambda\psi}),$$

which contradicts (9.3). Hence $(p_{\lambda\psi}, u_{\lambda\psi})$ is an upper solution, and it is either bounded or unbounded. If it is bounded, then $u_{\lambda\psi}(r) < 0$ for all $0 < r \leq \tilde{R}_{\lambda\psi}$, which is a contradiction because $u_{\lambda\psi}(r) > u_{\lambda\bar{\psi}}(r) = 0$ at $r = \tilde{R}_{\lambda\bar{\psi}}$. Hence $(p_{\lambda\psi}, u_{\lambda\psi})$ is an unbounded upper solution. Consider next the case $\psi < \bar{\psi}$. By assertion (2), $(p_{\lambda\psi}, u_{\lambda\psi})$ cannot be an upper solution, and by the result we have just proved, $(p_{\lambda\psi}, u_{\lambda\psi})$ cannot be a subsolution. It follows that $(p_{\lambda\psi}, u_{\lambda\psi})$ must be a lower solution.

Finally, assertion (3) follows immediately from (1) and (2). \square

Lemma 9.3. *For any $\lambda \in (0, \lambda_\infty)$, the set*

$$\Omega_1(\lambda) = \{\psi \in \mathbf{R} : (p_{\lambda\psi}, u_{\lambda\psi}) \text{ is an upper solution}\}$$

is bounded from below.

Proof. If the assertion is not true, then, by Lemma 9.2 (2), for any $\psi \in \mathbf{R}$, $(p_{\lambda\psi}, u_{\lambda\psi})$ is an upper solution. Using Lemma 9.1, we infer that, as ψ decreases, both $p_{\lambda\psi}(r)$ and $u_{\lambda\psi}(r)$ decrease, whereas the interval of definition of $(p_{\lambda\psi}, u_{\lambda\psi})$ increases (for unbounded upper solutions). Since $p_{\lambda\psi}(r)$ ($\psi \in \mathbf{R}$) are all positive and, by (8.1),

$$u_{\lambda\psi}(r) \geq C > -\infty, \quad C \text{ independent of } \psi,$$

the limits

$$p = \lim_{\psi \rightarrow -\infty} p_{\lambda\psi}, \quad u = \lim_{\psi \rightarrow -\infty} u_{\lambda\psi}$$

exist and, furthermore,

$$(9.4) \quad 0 \leq p < p_{\lambda\psi}, \quad u < u_{\lambda\psi} \quad \text{for all } \psi \in \mathbf{R}.$$

On the other hand, replacing (p, u) in (8.1), (8.2) with $(p_{\lambda\psi}, u_{\lambda\psi})$ and letting $\psi \rightarrow -\infty$, we see that (p, u) is a weak solution of (2.19)–(2.21), and, by Theorem 8.1, (p, u) is also a classical solution. It follows that there exists a $\psi_0 \in \mathbf{R}$ such that $p = p_{\lambda\psi_0}$ and $u = u_{\lambda\psi_0}$, which contradicts (9.4). Hence the desired assertion follows. \square

Lemma 9.4. *For any $\lambda \in (0, \lambda_\infty)$, the set*

$$\Omega_2(\lambda) = \{\psi \in \mathbf{R} : (p_{\lambda\psi}, u_{\lambda\psi}) \text{ is a lower solution}\}$$

is bounded from above.

Proof. If the assertion is false, then, by Lemma 9.2 (3), $(p_{\lambda\psi}, u_{\lambda\psi})$ is a lower solution for any $\psi \in \mathbf{R}$. By Theorem 7.4 (3), the functions $\{p_{\lambda\psi}(r) : \psi \in \mathbf{R}\}$ form a bounded subset of $L^\infty[0, R_\lambda]$. We can therefore find an unbounded increasing sequence $\{\psi_n\}_{n=1}^\infty$ such that the corresponding sequence $\{p_{\lambda\psi_n}\}_{n=1}^\infty$ is $*$ -weakly convergent in $L^\infty[0, R_\lambda]$. Let p denote the limit function. By (8.1), the corresponding sequence $\{u_{\lambda\psi_n}\}_{n=1}^\infty$ is uniformly convergent on $[0, R_\lambda]$. Let u denote the limit function. Replacing (p, u) in (8.1), (8.2) with $(p_{\lambda\psi_n}, u_{\lambda\psi_n})$ and letting $n \rightarrow \infty$, one easily finds that (p, u) is a weak solution of (2.19)–(2.21). We claim that there exists a $\bar{\psi} \in \mathbf{R}$ such that, for any $\psi > \bar{\psi}$,

$$(9.5) \quad \frac{d}{dr} p_{\lambda\psi}(r) > 0 \quad \text{for } 0 < r < \delta_\psi$$

for some $\delta_\psi > 0$. Indeed, if $\sigma(\lambda) > 1$, then by (5.43), (5.45) we have

$$p(r) = \alpha(\lambda) + \frac{1}{2}\alpha_2(\lambda)r^2 + o(r^2) \quad \text{as } r \rightarrow 0.$$

Since, by (4.10),

$$\alpha_2(\lambda) = \frac{1}{\gamma_2(\lambda)} \{K'_P(\lambda) + (K'_M(\lambda) - K'_N(\lambda)) - K'_M(\lambda)\alpha^2(\lambda)\} c''_\lambda(\lambda) > 0,$$

we see that (9.5) actually holds for all $\psi \in \mathbf{R}$. If $\sigma(\lambda) = 1$, then by (5.45) (taking $n = 1$) we see that (9.5) also holds for all $\psi \in \mathbf{R}$, because by (5.22),

$$\mu_0(\lambda) = \frac{1}{2\beta(\lambda)} \{K'_P(\lambda) + (K'_M(\lambda) - K'_N(\lambda)) - K'_M(\lambda)\alpha^2(\lambda)\} c''_\lambda(\lambda) < 0.$$

Finally, if $-1 < \sigma(\lambda) < 1$, then (9.5) follows from (5.43), (5.54) and (5.55), provided ψ is sufficiently large.

Having proved (9.5), we can now apply Lemma 9.1 to deduce that, if $\psi' > \psi > \bar{\psi}$, then

$$p_{\lambda\psi'}(r) > p_{\lambda\psi}(r), \quad u_{\lambda\psi'}(r) > u_{\lambda\psi}(r)$$

as long as $\frac{d}{dr}p_{\lambda\psi}(r) > 0$. Assuming for simplicity that $\psi_1 > \bar{\psi}$, we deduce, in particular, that

$$p_{\lambda\psi_n}(r) \geq p_{\lambda\psi_1}(r) \quad \text{for all } 0 < r < \tilde{R} \text{ and } n \geq 1,$$

where \tilde{R} is the first positive zero of $dp_{\lambda\psi_1}/dr$. It follows that

$$p(r) \geq p_{\lambda\psi_1}(r) \geq 0 \quad \text{for } 0 \leq r \leq \tilde{R},$$

so that, by Theorem 8.1, (p, u) is a classical solution on the interval $[0, \tilde{R}]$. By uniqueness of the solutions, we infer that there exists $\psi_0 \in \mathbf{R}$ so that $(p, u) = (p_{\lambda\psi_0}, u_{\lambda\psi_0})$. Now take a positive integer N large enough so that $\psi_N > \psi_0$, which ensures that, for some $\delta > 0$,

$$p_{\lambda\psi_N}(r) > p_{\lambda\psi_0}(r) \quad \text{for } 0 < r < \delta.$$

But since

$$p_{\lambda\psi_n}(r) \geq p_{\lambda\psi_N}(r) \quad \text{for } 0 < r < \tilde{R}' \text{ and } n \geq N,$$

where \tilde{R}' is the first positive zero of $dp_{\lambda\psi_N}/dr$, we have, by taking the $*$ -weak limit of the $p_{\lambda\psi_n}$,

$$p_{\lambda\psi_0}(r) \geq p_{\lambda\psi_N}(r) \quad \text{for } 0 \leq r \leq \tilde{R}',$$

which is a contradiction. This completes the proof of the lemma. \square

Remark 9.1. Since $p_{\lambda\psi}(r)$ is uniformly bounded for absolutely large negative ψ , one can find a sequence $\psi_m \rightarrow -\infty$ such that $p_{\lambda\psi_m}(r)$ $*$ -weakly converges in $L^\infty[0, R_\lambda]$ and $u_{\lambda\psi_m}(r)$ uniformly converges for $0 \leq r \leq R_\lambda$. The limits $p(r)$, $u(r)$ form a weak solution of (2.19)–(2.21). By similar arguments as in the above proof, one can show that this solution cannot be a classical solution such that $p(0) = \alpha(\lambda)$. Thus $p(0) = \alpha_-(\lambda)$, where $\alpha_-(\lambda)$ is the negative root of (3.1) (see (7.9)). One can further deduce that $p(r) \leq 0$ for all $0 \leq r \leq R_\lambda$.

Lemma 9.5. *There exists a $\bar{\lambda} \in (0, \lambda_\infty)$ such that for all $\lambda \in (\bar{\lambda}, \lambda_\infty)$ the set*

$$\Omega_3(\lambda) = \{\psi \in \mathbf{R} : (p_{\lambda\psi}, u_{\lambda\psi}) \text{ is a bounded upper solution}\}$$

is empty.

Proof. For any $\lambda \in (0, 1)$ we introduce the function

$$w_\lambda(r) = \frac{1}{r^2} \int_0^r \left(-K_D(c_\lambda(\rho)) + K_M(c_\lambda(\rho))\alpha(c_\lambda(\rho)) \right) \rho^2 d\rho, \quad 0 < r \leq R_\lambda,$$

and set $w_\lambda(0) = 0$. It is clear that w_λ depends continuously on λ in the topology of $C[0, R_\lambda]$. Clearly,

$$\frac{d}{dr}(r^2 w_\lambda(r)) = 3r^2 \beta(c_\lambda(r)), \quad 0 \leq r \leq R_\lambda,$$

where $\beta(\cdot)$ is the function defined in (3.6), so that by Lemma 3.2,

$$w_\lambda(r) > 0 \quad \text{for } 0 < r \leq R_\lambda, \quad \lambda_\infty \leq \lambda < 1,$$

and

$$w'_\lambda(0) = \beta(\lambda) < 0,$$

$$\frac{d}{dr}(w'_\lambda(r) + \frac{2}{r}w_\lambda(r)) > 0, \quad 0 \leq r \leq R_\lambda,$$

for $0 < \lambda < \lambda_\infty$. Hence there exists a $\bar{\lambda} \in (0, \lambda_\infty)$ such that for $\lambda \in (\bar{\lambda}, \lambda_\infty)$,

$$w_\lambda(r) < 0 \quad \text{for } 0 < r < \hat{R}_\lambda, \quad w_\lambda(r) > 0 \quad \text{for } \hat{R}_\lambda < r \leq R_\lambda,$$

for some point \hat{R}_λ . We claim that, for $\lambda \in (\bar{\lambda}, \lambda_\infty)$, the set $\Omega_3(\lambda)$ is empty. Indeed, if this is not the case, then there exists a bounded upper solution (p, u) for this λ , so that $p(r) > \alpha(c_\lambda(r))$ for all $0 < r \leq R_\lambda$. By (8.1), we then have $u(r) > w_\lambda(r)$ for $0 < r \leq R_\lambda$. Therefore, $u(r) > 0$ for $\hat{R}_\lambda \leq r \leq R_\lambda$. Since $u'(0) = \beta(\lambda) < 0$, so that $u(r) < 0$ for r near 0, we see that (p, u) is a subsolution, which is a contradiction. \square

Lemma 9.6. *Given $\lambda \in (0, \lambda_\infty)$, there exists an entire solution of the problem (2.19)–(2.21) which is a subsolution if and only if the set $\Omega_3(\lambda)$ is empty.*

Proof. If (2.19)–(2.21) has a subsolution, then by Lemma 9.2 (1), the set $\Omega_3(\lambda)$ is empty. Suppose conversely that $\Omega_3(\lambda)$ is empty. By Lemmas 9.3 and 9.4, $\Omega_1(\lambda)$ and $\Omega_2(\lambda)$ are bounded, respectively, from below and from above, and by Lemma 9.2 (1) there is at most one ψ such that $(p_{\lambda\psi}, u_{\lambda\psi})$ is a subsolution. Consequently, $\Omega_1(\lambda)$ and $\Omega_2(\lambda)$ are both nonempty and, by Lemma 9.2,

$$\inf \Omega_1(\lambda) = \sup \Omega_2(\lambda) \equiv \bar{\psi}.$$

We claim that $(p, u) \equiv (p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution. Indeed, using Theorem 6.1, we easily see that the set $\Omega_2(\lambda)$ is open, so that $\bar{\psi} \notin \Omega_2(\lambda)$. To prove that $\bar{\psi} \notin \Omega_1(\lambda)$ we assume the converse: $\bar{\psi} \in \Omega_1(\lambda)$. Let $[0, R)$ be the domain of definition of (p, u) , so that

$$\lim_{r \rightarrow R} p(r) = +\infty.$$

By the uniform boundedness of $\{p_{\lambda\psi} : \psi \in \Omega_2(\lambda)\}$, we can find an increasing sequence $\{\psi_n\}_{n=1}^\infty$ converging to $\bar{\psi}$, such that $p_{\lambda\psi_n}$ converges $*$ -weakly in $L^\infty[0, R_\lambda]$ to some function $\bar{p} \in L^\infty[0, R_\lambda]$, and $u_{\lambda\psi_n}$ converges uniformly to some function $\bar{u} \in C[0, R_\lambda]$. Replacing (p, u) in (8.1), (8.2) with $(p_{\lambda\psi_n}, u_{\lambda\psi_n})$ and letting $n \rightarrow \infty$, one easily finds that (\bar{p}, \bar{u}) is a weak solution of (2.19)–(2.21). Since $\lim_{n \rightarrow \infty} \psi_n = \bar{\psi}$, we conclude, as in the proof of Lemma 9.4, that $(p(r), u(r)) = (\bar{p}(r), \bar{u}(r))$ for r near 0, so that, by Theorem 8.3,

$$p(r) = \bar{p}(r), \quad u(r) = \bar{u}(r) \quad \text{for } 0 < r < R.$$

It follows that $p(r)$ is bounded for $r \in [0, R)$, which is a contradiction. Therefore, $\bar{\psi} \notin \Omega_1(\lambda)$.

Since we have proved that $\bar{\psi}$ does not belong to $\Omega_1(\lambda)$ nor to $\Omega_2(\lambda)$, it follows that (p, u) must be a subsolution. \square

By Lemmas 9.5 and 9.6, we immediately get the following result:

Lemma 9.7. *There exists a $\bar{\lambda} \in (0, \lambda_\infty)$ such that for any $\lambda \in (\bar{\lambda}, \lambda_\infty)$, there is a unique number $\bar{\psi} = \bar{\psi}(\lambda)$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution.*

10. EXISTENCE OF SOLUTIONS OF THE FREE BOUNDARY PROBLEM

In this section we prove the existence of a solution of the free boundary problem (2.19)–(2.22). If for some $\lambda \in (0, \lambda_\infty)$ there is a $\bar{\psi} \in \mathbf{R}$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a subsolution (resp. supersolution), then we say that λ is a *subsolution point* (resp. *supersolution point*). Similarly, if for some $\lambda \in (0, \lambda_\infty)$ there exists $\bar{\psi} \in \mathbf{R}$ such that $(p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}}, R_\lambda)$ is a solution of the free boundary problem (2.19)–(2.22), then we say that λ is a *solution point*. A subsolution that is not a solution of the free boundary problem will be called a *strict subsolution*, and the corresponding λ will be called a *strict subsolution point*. We introduce the sets

$$B_1 = \{\lambda \in (0, \lambda_\infty) : \lambda \text{ is a strict subsolution point}\},$$

$$\tilde{B}_1 = \{\lambda \in (0, \lambda_\infty) : \lambda \text{ is a subsolution point}\},$$

$$B_2 = \{\lambda \in (0, \lambda_\infty) : \lambda \text{ is a supersolution point}\}.$$

By Lemma 9.7, the set \tilde{B}_1 is nonempty; in fact, the proof of Lemma 9.5 shows that B_1 is nonempty. Later on we shall prove that B_2 is nonempty.

Lemma 10.1. *λ is a supersolution point if and only if it is not a subsolution point; in other words,*

$$(10.1) \quad \tilde{B}_1 \cup B_2 = (0, \lambda_\infty), \quad \tilde{B}_1 \cap B_2 = \emptyset.$$

Proof. Lemma 9.2 (1) shows that if λ is a subsolution point, then λ cannot be a supersolution point. To prove the converse, we note that if λ is not a subsolution point, then the set $\Omega_3(\lambda)$ is nonempty (by Lemma 9.6) and is bounded below (by Lemma 9.3). Let

$$\bar{\psi} = \inf \Omega_3(\lambda).$$

We claim that $(p, u) = (p_{\lambda\bar{\psi}}, u_{\lambda\bar{\psi}})$ is a supersolution. Indeed, using Lemma 9.1 and Theorem 8.1, one readily finds that p and u are the monotone decreasing limits of $p_{\lambda\psi}$ and $u_{\lambda\psi}$, respectively, as $\psi \rightarrow \bar{\psi} + 0$, $\psi \in \Omega_3(\lambda)$, so that they satisfy

$$p(r) > \alpha(c_\lambda(r)), \quad p'(r) > 0, \quad u(r) < 0 \quad \text{for } 0 < r < R_\lambda,$$

and $p(R_\lambda) \geq \alpha(c_\lambda(R_\lambda)) = 1$, $u(R_\lambda) < 0$. If $p(R_\lambda) > 1$, then for sufficiently small $\delta > 0$ we can find a corresponding $\varepsilon > 0$ such that

$$p(r) - \alpha(c_\lambda(r)) \geq \varepsilon, \quad u(r) \leq -\varepsilon$$

for all $\delta \leq r \leq R_\lambda$. By Theorem 6.1 and the standard ODE theory, it follows that if ψ is sufficiently close to $\bar{\psi}$, then

$$p_{\lambda\psi}(r) - \alpha(c_\lambda(r)) \geq \frac{1}{2}\varepsilon, \quad u_{\lambda\psi}(r) \leq -\frac{1}{2}\varepsilon$$

for $\delta \leq r \leq R_\lambda$, so that if also $\psi < \bar{\psi}$, then $\psi \in \Omega_3(\lambda)$, which is a contradiction. Hence $p(R_\lambda) = 1$, and (p, u) is a supersolution. \square

In the sequel we need to consider limits of sequences of functions $(p_{\lambda_n\psi}, u_{\lambda_n\psi})$ ($n = 1, 2, \dots$). Since these functions are not defined on a common interval, it will be convenient to make a transformation of variables $(r, p, u, c_\lambda) \rightarrow (\bar{r}, \bar{p}, \bar{u}, \bar{c}_\lambda)$:

$$\bar{r} = \frac{r}{R_\lambda}, \quad \bar{u}(\bar{r}) = \frac{u(\bar{r}R_\lambda)}{R_\lambda}, \quad \bar{p}(\bar{r}) = p(\bar{r}R_\lambda), \quad \bar{c}_\lambda(\bar{r}) = c_\lambda(\bar{r}R_\lambda).$$

It is easy to verify that equations (2.19)–(2.21) are invariant under this change of variables, but the interval $[0, R_\lambda]$ is changed into the unit interval $[0, 1]$. Clearly, all

results established in previous sections are still valid for the transformed problem. In the sequel we shall always write, for brevity, the variables \bar{r} , \bar{p} , \bar{u} and \bar{c}_λ as respectively r , p , u and c_λ . It should be noted that the values $p(1)$, $u(1)$ in the new variables are respectively equal to the values $p(R_\lambda)$ and $u(R_\lambda)/R_\lambda$ in the old variables.

Lemma 10.2. *The sets B_1 , B_2 are open.*

Proof. To prove that B_1 is open we assume the converse, that is, there exists a point $\tilde{\lambda} \in B_1$ that is the limit of a sequence of points $\{\lambda_m\}_{m=1}^\infty$ that are not strict subsolution points. By Lemma 10.1, for each λ_m we have a corresponding (p_m, u_m) that is either a supersolution or a solution of the free boundary problem. In both cases we have, for each m ,

$$\alpha(c_{\lambda_m}(r)) \leq p_m(r) \leq 1, \quad p'_m(r) \geq 0, \quad u_m(r) \leq 0 \quad \text{for } 0 \leq r \leq 1.$$

It follows that $\{p_m\}_{m=1}^\infty$ has a subsequence that is $*$ -weakly convergent in $L^\infty[0, 1]$. For simplicity we assume that $\{p_m\}_{m=1}^\infty$ is such a subsequence, and denote by p the limit function. Then $\{u_m\}_{m=1}^\infty$ converges uniformly to a function $u \in C[0, 1]$. Clearly,

$$\alpha(c_{\tilde{\lambda}}(r)) \leq p(r) \leq 1, \quad u(r) \leq 0 \quad \text{for } 0 \leq r \leq 1,$$

and $p(r)$ is monotone nondecreasing for $0 \leq r \leq 1$. On the other hand, replacing λ , p and u in (8.1), (8.2) with, respectively, λ_m , p_m and u_m and letting $m \rightarrow \infty$, we see that (p, u) is a weak solution of (2.19)–(2.21) with respect to $\lambda = \tilde{\lambda}$. Since $p \geq 0$, (p, u) is a classical solution, by Theorem 8.1, so that the above properties of (p, u) imply that

$$\alpha(c_{\tilde{\lambda}}(r)) < p(r) < 1, \quad p'(r) > 0, \quad u(r) < 0 \quad \text{for } 0 < r < 1,$$

and $p(1) = 1$. Furthermore, we have either $u(1) = 0$ or $u(1) < 0$. Clearly, in the first case (p, u) is a solution of the free boundary problem, and in the second case (p, u) is a supersolution, so that in either case $\tilde{\lambda} \notin B_1$, which is a contradiction. Hence the set B_1 is open.

Next we prove that B_2 is open. Suppose that $\tilde{\lambda} \in B_2$ and let $(\tilde{p}_{\tilde{\lambda}}, \tilde{u}_{\tilde{\lambda}}) = (p_{\tilde{\lambda}\tilde{\psi}}, u_{\tilde{\lambda}\tilde{\psi}})$ be the corresponding supersolution. Then we have

$$\tilde{p}_{\tilde{\lambda}}(r) > \alpha(c_{\tilde{\lambda}}(r)), \quad \tilde{p}'_{\tilde{\lambda}}(r) > 0, \quad \tilde{u}_{\tilde{\lambda}}(r) < 0 \quad \text{for } 0 < r < 1,$$

and $\tilde{p}_{\tilde{\lambda}}(1) = 1$, $\tilde{u}_{\tilde{\lambda}}(1) < 0$. It follows by Theorem 6.1 and Lemma 9.1 that we can find a $\hat{\psi} > \tilde{\psi}$ sufficiently near $\tilde{\psi}$ such that

$$p_{\tilde{\lambda}\hat{\psi}}(r) > \alpha(c_{\tilde{\lambda}}(r)), \quad u_{\tilde{\lambda}\hat{\psi}}(r) < 0 \quad \text{for } 0 < r \leq 1$$

(and consequently also $p'_{\tilde{\lambda}\hat{\psi}}(r) > 0$ for all $0 < r \leq 1$). This implies that for any $\delta > 0$ sufficiently small, there exists a corresponding $\varepsilon > 0$ such that

$$p_{\tilde{\lambda}\hat{\psi}}(r) - \alpha(c_{\tilde{\lambda}}(r)) \geq \varepsilon, \quad u_{\tilde{\lambda}\hat{\psi}}(r) \leq -\varepsilon \quad \text{for } \delta \leq r \leq 1.$$

By Theorem 6.1, it follows that for (λ, ψ) sufficiently close to $(\tilde{\lambda}, \hat{\psi})$,

$$p_{\lambda\psi}(r) - \alpha(c_\lambda(r)) \geq \tfrac{1}{2}\varepsilon, \quad u_{\lambda\psi}(r) \leq -\tfrac{1}{2}\varepsilon \quad \text{for } \delta \leq r \leq 1,$$

which implies that $(p_{\lambda\psi}, u_{\lambda\psi})$ is a bounded upper solution. Hence for λ in a small neighborhood of $\tilde{\lambda}$ the set $\Omega_3(\lambda)$ is nonempty and, therefore, by Lemma 9.6, these λ 's belong to B_2 , so that B_2 is open. □

When $\lambda = 0$, the problem (2.19)–(2.21) (in the new variables) becomes

$$(10.2) \quad u'(r) + \frac{2}{r}u(r) = -K_D(0)(1 - p(r)), \quad 0 < r < 1,$$

$$(10.3) \quad u(0) = 0,$$

$$(10.4) \quad u(r)p'(r) = -\{(K_Q(0) - K_D(0)) + K_D(0)p(r)\}p(r), \quad 0 < r < 1.$$

The set of all solutions of (10.2)–(10.4) is characterized by the following lemma.

Lemma 10.3. *For every $\omega \in \mathbf{R}$, (10.2)–(10.4) has a unique solution satisfying*

$$(10.5) \quad \begin{aligned} p(r) &= \alpha(0) + \omega r^{1+\sigma(0)} + o(r^{1+\sigma(0)}), \\ u(r) &= \beta(0)r + \frac{\omega K_D(0)}{4 + \sigma(0)} r^{2+\sigma(0)} + o(r^{2+\sigma(0)}) \end{aligned}$$

as $r \rightarrow 0$; furthermore, (i) if $\omega = 0$, then

$$p(r) = \alpha(0), \quad u(r) = \beta(0)r$$

for all $0 \leq r \leq 1$; (ii) if $\omega > 0$, then (p, u) is an upper solution; (iii) if $\omega < 0$, then (p, u) is a lower solution.

Proof. The formula (10.5) and the uniqueness follow as in the proofs of Theorems 5.3 and 5.4. By direct computation one finds that $(p_0(r), u_0(r)) = (\alpha(0), \beta(0)r)$ is a solution of (10.2)–(10.4), so that there exists $\omega \in \mathbf{R}$ such that it coincides with the solution given by (10.5), and clearly $(p, u) = (p_0, u_0)$ for $\omega = 0$. Next, noticing that $1 + \sigma(0) > 0$, we see that $p'(r) < 0$ for r near 0 if $\omega < 0$, so that (p, u) is a lower solution if $\omega < 0$.

Consider next the case $\omega > 0$. Since $p'(r) = (1 + \sigma(0))\omega r^{\sigma(0)} + o(r^{\sigma(0)})$ (as $r \rightarrow 0$) and $u(0) = \beta(0) < 0$, we can find a number $\delta > 0$ such that

$$(10.6) \quad p'(r) > 0, \quad u(r) < 0$$

for $0 < r < \delta$. Let $(0, R)$ be the largest open interval such that (10.6) holds for $r \in (0, R)$. Then either $\lim_{r \rightarrow R} p(r) = +\infty$ or $\lim_{r \rightarrow R} p(r) = a$ for some $\alpha(0) < a < \infty$. In the first case (p, u) is clearly an unbounded upper solution. In the second case the right-hand side of (10.2) is bounded, so that $b \equiv \lim_{r \rightarrow R} u(r)$ exists, and clearly $b \leq 0$. If $b < 0$ and $R < 1$, then we can extend (p, u) to a larger interval such that $u(r) < 0$ in this interval. Also $p'(r) > 0$ in this interval, for otherwise, by (10.4), $p(\bar{r}) = \alpha(0)$ at the first point \bar{r} where $p'(\bar{r}) = 0$, which is a contradiction since $p(0) = \alpha(0)$ and $p(r)$ is strictly increasing in $0 \leq r \leq \bar{r}$. Hence (10.6) still holds on this interval, which is contrary to the maximality of $(0, R)$. Thus either $b < 0$ and $R = 1$, or $b = 0$. In the first case (p, u) is a bounded upper solution. If we prove that the second case cannot occur, then the proof that (p, u) is an upper solution (when $\omega > 0$) will be completed. Suppose $b = 0$. Then by (8.1), (8.2) and l'Hospital's rule,

$$\begin{aligned} a &= \lim_{r \rightarrow R} p(r) = \lim_{r \rightarrow R} \frac{\int_0^r K_Q(0)p(\rho)\rho^2 d\rho}{\int_0^r K_D(0)(1-p(\rho))\rho^2 d\rho} \\ &= \lim_{r \rightarrow R} \frac{K_Q(0)p(r)}{K_D(0)(1-p(r))} = \frac{K_Q(0)a}{K_D(0)(1-a)}, \end{aligned}$$

so that $a = \alpha(0)$, which is a contradiction. \square

Lemma 10.4. *The set B_2 is nonempty. More precisely, there exists a $\hat{\lambda} \in (0, 1)$ such that $(0, \hat{\lambda}) \subset B_2$.*

Proof. If the assertion is not true, then we can find a monotone decreasing sequence $\{\lambda_m\}_{m=1}^\infty$ converging to 0 such that each λ_m is a subsolution point. Let (p_m, u_m) be the corresponding sequence of subsolutions. Since $\{p_m\}_{m=1}^\infty$ is uniformly bounded, we may assume that p_m *-weakly converges to some function $p \in L^\infty[0, 1]$. It follows that u_m uniformly converges to some function $u \in C[0, 1]$. By a similar argument as before, we easily infer that (p, u) is a solution of the problem (10.2)–(10.4) with $p(r)$ monotone increasing. Using Lemma 10.3, we conclude that (p, u) either is a bounded upper solution or is equal to (p_0, u_0) . It follows that $u(1) < 0$. On the other hand, since (p_m, u_m) are subsolutions, we have $u_m(1) \geq 0$, so that also $u(1) \geq 0$, which is a contradiction. \square

Proof of Theorem 2.1. The sets B_1, B_2 are both open (by Lemma 10.2), nonempty (by Lemmas 9.7 and 10.4), and they are disjoint (by Lemma 10.1). Hence

$$(0, \lambda_\infty) \neq B_1 \cup B_2.$$

Recalling (10.1), we conclude that $\tilde{B}_1 \setminus B_2 \neq \emptyset$, so that there exists at least one solution point. \square

11. UNIQUENESS OF THE SOLUTION

In this section we prove the uniqueness of the solution of the free boundary problem (2.1)–(2.6). We assume that there exist two different solutions (c_1, p_1, u_1, R_1) and (c_2, p_2, u_2, R_2) , and derive a contradiction.

By Lemma 9.2, $R_1 \neq R_2$, and we may take $R_1 < R_2$. Introducing the functions

$$\bar{c}_i(r) = c_i(rR_i), \quad \bar{p}_i(r) = p_i(rR_i), \quad \bar{u}_i(r) = \frac{u_i(rR_i)}{R_i}, \quad 0 \leq r \leq 1, \quad i = 1, 2,$$

we find that the $(\bar{c}_i, \bar{p}_i, \bar{u}_i)$'s satisfy the system of equations

$$(11.1) \quad \bar{c}_i''(r) + \frac{2}{r}\bar{c}_i'(r) = R_i^2 F(\bar{c}_i(r)), \quad 0 < r < 1,$$

$$(11.2) \quad \bar{c}_i'(0) = 0, \quad \bar{c}_i(1) = 1,$$

$$(11.3) \quad \bar{u}_i'(r) + \frac{2}{r}\bar{u}_i(r) = -K_D(\bar{c}_1(r)) + K_M(\bar{c}_i(r))\bar{p}_i(r), \quad 0 < r < 1,$$

$$(11.4) \quad \bar{u}_i(0) = 0,$$

$$(11.5) \quad \bar{u}_i(r)\bar{p}_i'(r) = K_P(\bar{c}_i(r)) + (K_M(\bar{c}_i(r)) - K_N(\bar{c}_i(r)))\bar{p}_i(r) - K_M(\bar{c}_i(r))\bar{p}_i^2(r), \quad 0 < r < 1,$$

$$(11.6) \quad \bar{u}_i(1) = 0.$$

Since $R_1 < R_2$, from (11.1) and (11.2) we get, by comparison, that

$$(11.7) \quad \bar{c}_1(r) > \bar{c}_2(r) \quad \text{for } 0 \leq r < 1$$

and, by the maximum principle, that

$$(11.8) \quad \bar{c}_1'(1) < \bar{c}_2'(1).$$

Clearly

$$(11.9) \quad \bar{p}_1(1) = \bar{p}_2(1) = 1,$$

so that, by (11.3),

$$(11.10) \quad \bar{u}'_1(1) = \bar{u}'_2(1) = K_B(1)$$

and, by (7.4),

$$\begin{aligned} \bar{p}'_i(1) &= \frac{\{K'_P(1) + (K'_M(1) - K'_N(1))\bar{p}_i(1) - K'_M(1)\bar{p}_i^2(1)\}\bar{c}'_i(1)}{\bar{u}'_i(1) + 2K_M(1)\bar{p}_i(1) - (K_M(1) - K_N(1))} \\ &= \frac{-K'_Q(1)\bar{c}'_i(1)}{2K_B(1) + K_P(1)} \equiv A\bar{c}'_i(1), \end{aligned}$$

where $A > 0$. Recalling (11.8), we conclude that

$$(11.11) \quad \bar{p}'_1(1) < \bar{p}'_2(1).$$

From (11.9) and (11.11) it follows that there exists a $\delta > 0$ such that

$$\bar{p}_1(r) > \bar{p}_2(r) \quad \text{for } 1 - \delta < r < 1.$$

Let $(r_0, 1)$ be the largest open interval such that

$$(11.12) \quad \bar{p}_1(r) > \bar{p}_2(r) \quad \text{for } r_0 < r < 1; \quad 0 \leq r_0 < 1.$$

By (11.3) we have

$$(11.13) \quad \bar{u}_i(r) = \frac{1}{r^2} \int_0^r \{ -K_D(\bar{c}_i(\rho)) + K_M(\bar{c}_i(\rho))\bar{p}_i(\rho) \} \rho^2 d\rho, \quad 0 < r \leq 1.$$

Since the function $-K_D(c) + K_M(c)\alpha$ is strictly monotone increasing in both c and α , using (11.7) and (11.12) we infer that

$$(11.14) \quad -K_D(\bar{c}_1(\rho)) + K_M(\bar{c}_1(\rho))\bar{p}_1(\rho) > -K_D(\bar{c}_2(\rho)) + K_M(\bar{c}_2(\rho))\bar{p}_2(\rho) \quad \text{for } r_0 < r < 1.$$

Thus, if $r_0 = 0$, then, by (11.13), $\bar{u}_1(1) > \bar{u}_2(1)$, which is a contradiction to (11.6). It follows that $r_0 > 0$, and then

$$(11.15) \quad \bar{p}_1(r_0) = \bar{p}_2(r_0),$$

which implies, by (11.12), that

$$(11.16) \quad \bar{p}'_1(r_0) \geq \bar{p}'_2(r_0).$$

Since

$$\int_0^1 \{ -K_D(\bar{c}_i(\rho)) + K_M(\bar{c}_i(\rho))\bar{p}_i(\rho) \} \rho^2 d\rho = \bar{u}_i(1) = 0,$$

(11.13) gives

$$\bar{u}_i(r_0) = -\frac{1}{r_0^2} \int_{r_0}^1 \{ -K_D(\bar{c}_i(\rho)) + K_M(\bar{c}_i(\rho))\bar{p}_i(\rho) \} \rho^2 d\rho,$$

so that, by (11.14),

$$\bar{u}_1(r_0) < \bar{u}_2(r_0).$$

Combining this with (11.16) and recalling that $\bar{u}_i < 0$, $\bar{p}'_i > 0$ (by Theorem 7.1), we conclude that

$$\bar{u}_1(r_0)\bar{p}'_1(r_0) < \bar{u}_2(r_0)\bar{p}'_2(r_0).$$

On the other hand, since the function

$$K_P(c) + (K_M(c) - K_N(c))\alpha - K_M(c)\alpha^2 = K_P(c)(1 - \alpha) + K_M(c)\alpha(1 - \alpha) - K_Q(c)\alpha$$

is strictly monotone increasing in c (for fixed $0 < \alpha < 1$), we infer from (11.5), (11.7) and (11.15) that

$$\bar{u}_1(r_0)\bar{p}'_1(r_0) > \bar{u}_2(r_0)\bar{p}'_2(r_0),$$

which is a contradiction. □

Remark 11.1. The proof of the existence part of Theorem 2.1 shows that if λ_1 is a supersolution point and λ_2 is a subsolution point, then there exists a solution point of the free boundary problem in the interval with endpoints λ_1, λ_2 . Furthermore, points λ_1 near 0 are supersolution points, and points λ_2 near λ_∞ ($\lambda_2 < \lambda_\infty$) are subsolution points. This fact, combined with the uniqueness part of Theorem 2.1, implies the following: If λ_* is the unique solution point of the free boundary problem, then every $\lambda \in (0, \lambda_*)$ is a supersolution point, and every $\lambda \in (\lambda_*, \lambda_\infty)$ is a subsolution point.

Remark 11.2. The assumption $K_D(1) = K_Q(1) = 0$ used throughout the paper can be weakened as follows:

$$K_D(1) \geq 0, \quad K_Q(1) \geq 0 \quad \text{and} \quad K_Q(1) < K_B(1).$$

In this case, $\alpha(1) \leq 1$, so that for a solution point λ , $u(R_\lambda) = 0$ and $p(R_\lambda) = \alpha(1) \leq 1$, rather than $p(R_\lambda) = 1$ as in Theorem 2.1. Other than this and other similar differences, all the results of §§4–11 hold with minor changes; for example, in Theorem 7.2 (a) one replaces $p(R) = 1$ by $p(R) \leq 1$, and in §11 one replaces (11.10) by

$$\bar{u}'_1(1) = \bar{u}'_2(1) = K_B(1)\alpha(1).$$

12. APPENDIX: THE PROOFS OF (5.30) AND (6.18)–(6.19)

Lemma 12.1. *Let $\bar{\lambda} \in (0, \lambda_\infty)$ and $\sigma(\bar{\lambda}) \geq n$, where n is an integer ≥ 2 . Let λ be a number in a small neighborhood of $\bar{\lambda}$. Assume that*

(12.1)
$$P(r) = \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-1} + v(r), \qquad U(r) = \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} + w(r).$$

Then

(12.2)
$$\begin{aligned} &(\beta(\lambda) + U(r)) \Big(f_\lambda(r, P(r), U(r)) - \sum_{m=0}^{n-1} \mu_m(\lambda) r^m \Big) \\ &= r^n y(r) + z_1(r) w(r) + r z_2(r) v(r) - K_M(c_\lambda(r)) v^2(r), \end{aligned}$$

where

(12.3)
$$|y(r)| \leq C, \qquad |z_1(r)| \leq C, \qquad |z_2(r)| \leq C,$$

and C is a constant independent of v, w and λ .

Proof. Let $p(r) = \alpha(\lambda) + rP(r)$. Since $\alpha_1(\lambda) = 0$, $\beta_1(\lambda) = \beta(\lambda)$ and $\beta_2(\lambda) = 0$, we have

$$\begin{aligned} p(r) &= \sum_{m=0}^n \frac{\alpha_m(\lambda)}{m!} r^m + r v(r) \equiv p_n(r) + r v(r), \\ \beta(\lambda) + U(r) &= \sum_{m=1}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} + w(r). \end{aligned}$$

By (5.11), (5.12), (5.13) and (5.8) we see that

$$\begin{aligned}
 & (\beta(\lambda) + U(r))f_\lambda(r, P(r), U(r)) \\
 &= -K_M(c_\lambda(r))P^2(r) - (1/r)g_\lambda(r)P(r) + (h_\lambda(r)/r^2) \\
 & \quad + (g_\lambda(0)/\beta(\lambda))(1/r)(\beta(\lambda) + U(r))P(r) \\
 &= (1/r^2)\{K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda))p(r) - K_M(c_\lambda(r))p^2(r)\} \\
 & \quad - (\sigma(\lambda) + 1)(1/r)(\beta(\lambda) + U(r))P(r) \\
 &\equiv I_1 - I_2.
 \end{aligned}$$

Clearly,

$$\begin{aligned}
 I_1 &= (1/r^2)\{K_P(c_\lambda(r)) + (K_M(c_\lambda(r)) - K_N(c_\lambda))p_n(r) - K_M(c_\lambda(r))p_n^2(r)\} \\
 & \quad + (1/r)\{(K_M(c_\lambda(r)) - K_N(c_\lambda)) \\
 & \quad - 2K_M(c_\lambda(r))p_n(r)v(r) - K_M(c_\lambda(r)) \cdot rv^2(r)\} \\
 &\equiv I_{11} + I_{12}.
 \end{aligned}$$

Using Taylor's expansions of the functions $K_P(c_\lambda(r))$, $K_M(c_\lambda(r))$ and $K_N(c_\lambda(r))$ up to order n about $r = 0$, and (5.19), (5.21), and recalling (3.1) and the relations

$$k_i^{(1)}(\lambda) = \frac{\partial}{\partial r} K_i(c_\lambda(r))|_{r=0} = K'_i(\lambda)c'_\lambda(0) = 0, \quad i = D, M, N, P,$$

we compute that

$$\begin{aligned}
 I_{11} &= \sum_{m=2}^{n+1} \frac{k_P^{(m)}(\lambda)}{m!} r^{m-2} + \left(\sum_{m=2}^{n+1} \frac{k_M^{(m)}(\lambda) - k_N^{(m)}(\lambda)}{m!} r^{m-2} \right) \left(\sum_{m=0}^n \frac{\alpha_m(\lambda)}{m!} r^m \right) \\
 & \quad - \left(\sum_{m=2}^{n+1} \frac{k_M^{(m)}(\lambda)}{m!} r^{m-2} \right) \left(\sum_{m=0}^n \frac{\alpha_m(\lambda)}{m!} r^m \right)^2 \\
 & \quad + (K_M(\lambda) - K_N(\lambda)) \left(\sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-2} \right) \\
 & \quad - K_M(\lambda) \left(2\alpha(\lambda) \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-2} \right. \\
 & \quad \left. + r^{-2} \left(\sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^m \right)^2 \right) + r^n y_1(r),
 \end{aligned}$$

or, by expanding the products,

$$\begin{aligned}
 I_{11} &= \sum_{m=0}^{n-1} \frac{r^m}{(m+2)!} \{ k_P^{m+2}(\lambda) + \sum_{j=2}^{m+2} \binom{m+2}{j} (k_M^{(j)}(\lambda) - k_N^{(j)}(\lambda)) \alpha_{m+2-j}(\lambda) \\
 & \quad - \sum_{j=2}^{m+2} \sum_{k=0}^{m-j+2} \binom{m+2}{j} \binom{m-j+2}{k} k_M^{(j)}(\lambda) \alpha_k(\lambda) \alpha_{m-j-k+2}(\lambda) \\
 & \quad - K_M(\lambda) \sum_{j=2}^m \binom{m+2}{j} \alpha_j(\lambda) \alpha_{m-j+2}(\lambda) \} \\
 & \quad + \left((K_M(\lambda) - K_N(\lambda)) - 2K_M(\lambda)\alpha(\lambda) \right) \left(\sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-2} \right) + r^n y_2(r),
 \end{aligned}$$

so that, by the relation

$$(12.4) \quad K_M(\lambda) - K_N(\lambda) - 2K_M(\lambda)\alpha(\lambda) = (\sigma(\lambda) + 1)\beta(\lambda)$$

(which follows from (3.8)) and the definition of μ_m (see (5.21) and (5.19)),

$$(12.5) \quad \begin{aligned} I_{11} &= \beta(\lambda) \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + \sum_{m=1}^{n-1} \frac{r^m}{(m+2)!} \sum_{j=2}^{m+1} \binom{m+2}{j} \beta_j(\lambda) \alpha_{m-j+3}(\lambda) \\ &\quad + (\sigma(\lambda) + 1)\beta(\lambda) \cdot \frac{1}{r} (P(r) - v(r)) + r^n y_3(r) \\ &\equiv J_1 + J_2 + J_3 + J_4, \end{aligned}$$

where $y_i(r)$ ($i = 1, 2, 3$) are functions depending only on $\alpha_m(\lambda)$ ($0 \leq m \leq n$), K_P , K_M , K_N and c_λ , so that they are uniformly bounded for all λ near $\bar{\lambda}$. Later on we shall also use $y_4(r)$, $z_1(r)$, $z_2(r)$ to denote various functions possessing similar properties. Since

$$\begin{aligned} &\left(\sum_{j=2}^{n+1} \frac{\beta_j(\lambda)}{j!} r^{j-1} \right) \left(\sum_{k=1}^{n-1} \frac{\alpha_{k+1}(\lambda)}{k!} r^{k-1} \right) \\ &= \left(\sum_{j=1}^n \frac{\beta_{j+1}(\lambda)}{(j+1)!} r^j \right) \left(\sum_{k=0}^{n-2} \frac{\alpha_{k+2}(\lambda)}{(k+1)!} r^k \right) \\ &= \sum_{j=1}^n \sum_{k=0}^{n-2} \frac{r^{j+k}}{(j+k+2)!} \cdot \frac{(j+k+2)!}{(j+1)!(k+1)!} \cdot \beta_{j+1}(\lambda) \alpha_{k+2}(\lambda) \\ &= \sum_{m=1}^{n-1} \frac{r^m}{(m+2)!} \sum_{j=2}^{m+1} \binom{m+2}{j} \beta_j(\lambda) \alpha_{m-j+3}(\lambda) + r^n y_4(r) \\ &= J_2 + r^n y_4(r), \end{aligned}$$

we have

$$\begin{aligned} J_2 &= \left(\sum_{m=2}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1} \right) \left(\sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{m!} r^{m-1} \right) - r^n y_4(r) \\ &= (U(r) - w(r)) \left(\sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{m!} r^{m-1} \right) - r^n y_4(r). \end{aligned}$$

Using (5.21), one may easily verify that

$$\frac{\alpha_{m+1}(\lambda)}{m!} = \mu_{m-1}(\lambda) + (\sigma(\lambda) + 1) \cdot \frac{\alpha_{m+1}(\lambda)}{(m+1)!}$$

($m = 1, 2, \dots, n-1$), so that

$$\sum_{m=1}^{n-1} \frac{\alpha_{m+1}(\lambda)}{m!} r^{m-1} = \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + (\sigma(\lambda) + 1) \cdot \frac{1}{r} (P(r) - v(r)) - \mu_{n-1}(\lambda) r^{n-1}.$$

Hence

$$\begin{aligned} J_2 &= (U(r) - w(r)) \left\{ \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + (\sigma(\lambda) + 1) \cdot \frac{1}{r} (P(r) - v(r)) \right\} \\ &\quad - \mu_{n-1}(\lambda) r^{n-1} (U(r) - w(r)) - r^n y_4(r). \end{aligned}$$

Substituting this into (12.5), we obtain

$$\begin{aligned}
 I_{11} &= \beta(\lambda) \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + (U(r) - w(r)) \cdot \sum_{m=0}^{n-1} \mu_m(\lambda) r^m \\
 &\quad + (\sigma(\lambda) + 1) \cdot (U(r) - w(r)) \cdot (1/r) (P(r) - v(r)) \\
 &\quad - \mu_{n-1}(\lambda) r^{n-1} (U(r) - w(r)) \\
 &\quad (\sigma(\lambda) + 1) \beta(\lambda) \cdot (1/r) (P(r) - v(r)) + r^n y_3(r) - r^n y_4(r) \\
 &= (U(r) + \beta(\lambda)) \cdot \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + (\sigma(\lambda) + 1) \cdot (1/r) (U(r) + \beta(\lambda)) P(r) \\
 &\quad + r^n y(r) + z_1(r) w(r) - (\sigma(\lambda) + 1) \beta \cdot (1/r) v(r),
 \end{aligned}$$

where

$$\begin{aligned}
 y(r) &= y_3(r) - y_4(r) - \mu_{n-1}(\lambda) r^{n-1} \cdot \sum_{m=3}^{n+1} \frac{\beta_m(\lambda)}{m!} r^{m-1}, \\
 z_1(r) &= - \sum_{m=0}^{n-1} \mu_m(\lambda) r^m - (\sigma(\lambda) + 1) \cdot \sum_{m=2}^n \frac{\alpha_m(\lambda)}{m!} r^{m-2}.
 \end{aligned}$$

Recalling (12.4) and noticing that

$$\left. \frac{d}{dr} \right|_{r=0} (K_M(c_\lambda(r)) - K_N(c_\lambda(r)) - 2K_M(c_\lambda(r))p_n(r)) = 0,$$

we easily find that

$$I_{12} - (\sigma(\lambda) + 1) \beta(\lambda) \cdot \frac{1}{r} v(r) = r z_2(r) v(r) - K_M(c_\lambda(r)) v^2(r).$$

Hence

$$\begin{aligned}
 (U(r) + \beta(\lambda)) f_\lambda(r, P(r), U(r)) &= I_{11} + I_{12} - I_2 \\
 &= (U(r) + \beta(\lambda)) \cdot \sum_{m=0}^{n-1} \mu_m(\lambda) r^m + r^n y(r) \\
 &\quad + z_1(r) w(r) + r z_2(r) v(r) - K_M(c_\lambda(r)) v^2(r),
 \end{aligned}$$

and the lemma follows. \square

Formulas (6.18) and (6.19) follow immediately from Lemma 11.1. To prove (5.30), recall that $\alpha_1(\lambda) = \beta_2(\lambda) = 0$, so that, by (5.28) and (5.29), (12.1) holds with $v(r)$ and $w(r)$ satisfying

$$|v(r)| \leq C r^n, \quad |w(r)| \leq C r^{n+1} \quad \text{for } 0 < r < \delta.$$

Hence the right-hand side of (12.2) is bounded by $C r^n$. Since also

$$|\beta(\lambda) + U(r)| \geq \frac{1}{2} |\beta(\lambda)| > 0 \quad \text{for } 0 < r < \delta,$$

we see that (5.30) follows. \square

ACKNOWLEDGEMENT

The first author is supported by the China National Natural Science Foundation (Grant 10171112) and the Distinguished Visiting Scholar Program of the China Scholarship Council. The second author is supported by the National Natural Science Foundation of USA Grant DMS-0098520. The first author wishes to thank the staff of the Department of Mathematics of The Ohio State University for their friendly hospitality when he was a visiting scholar from October, 2001 to September, 2002.

REFERENCES

- [1] J. Adam, A simplified mathematical model of tumor growth, *Math. Biosci.* 81(1986), 224–229.
- [2] N. Britton and M. Chaplain, A qualitative analysis of some models of tissue growth, *Math. Biosci.* 113(1993), 77–89.
- [3] H. Byrne and M. Chaplain, Growth of necrotic tumors in the presence and absence of inhibitors, *Math. Biosci.* 135(1996), 187–216.
- [4] S. Cui and A. Friedman, Analysis of a mathematical model of the effect of inhibitors on the growth of tumors, *Math. Biosci.* 164(2000), 103–137. MR **2001d**:92006
- [5] S. Cui and A. Friedman, Analysis of a mathematical model of the growth of necrotic tumors, *J. Math. Anal. Appl.* 255(2001), 636–677. MR **2002a**:35195
- [6] M. Dorie, R. Kallman and M. Coyne, Effect of cytochalasin b, nocodazole and irradiation on migration and internalization of cells and microspheres in tumor cell spheroids, *Exp. Cell Res.* 166(1986), 370–378.
- [7] M. Dorie, R. Kallman, D. Rapacchietta and *et al*, Migration and internalization of cells and polystyrene microspheres in tumor cell spheroids, *Exp. Cell Res.* 141(1982), 201–209.
- [8] A. Friedman and F. Reitich, Analysis of a mathematical model for the growth of tumors, *J. Math. Biol.* 38(1999), 262–284. MR **2001f**:92011
- [9] H. Greenspan, Models for the growth of solid tumor by diffusion, *Stud. Appl. Math.* 51(1972), 317–340.
- [10] H. Greenspan, On the growth and stability of cell cultures and solid tumors, *J. Theor. Biol.* 56(1976), 229–242. MR **55**:2183
- [11] F. Hughes and C. McCulloch, Quantification of chemotactic response of quiescent and proliferating fibroblasts in Boyden chambers by computer-assisted image analysis, *J. Histochem. Cytochem.* 39(1991), 243–246.
- [12] D. McElwain and G. Pettet, Cell migration in multicell spheroids: swimming against the tide, *Bull. Math. Biol.* 55(1993), 655–674.
- [13] G. Pettet, C. P. Please, M. J. Tindall and *et al*, The migration of cells in multicell tumor spheroids, *Bull. Math. Biol.* 63(2001), 231–257.
- [14] J. Sherratt and M. Chaplain, A new mathematical model for avascular tumor growth, *J. Math. Biol.* 43(2001), 291–312.
- [15] K. Thompson and H. Byrne, Modelling the internalisation of labelled cells in tumor spheroids, *Bull. Math. Biol.* 61(1999), 601–623.
- [16] J. Ward and J. King, Mathematical modelling of avascular-tumor growth II: Modelling growth saturation, *IMA J. Math. Appl. Med. Biol.* 15(1998), 1–42.

INSTITUTE OF MATHEMATICS, ZHONGSHAN UNIVERSITY, GUANGZHOU, GUANGDONG 510275, PEOPLE'S REPUBLIC OF CHINA

E-mail address: mcinst@zsu.edu.cn

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 231 WEST 18TH AVENUE, COLUMBUS, OHIO 43210-1174

E-mail address: afriedman@mbi.osu.edu

SHARP FOURIER TYPE AND COTYPE WITH RESPECT TO COMPACT SEMISIMPLE LIE GROUPS

JOSÉ GARCÍA-CUERVA, JOSÉ MANUEL MARCO, AND JAVIER PARCET

ABSTRACT. Sharp Fourier type and cotype of Lebesgue spaces and Schatten classes with respect to an arbitrary compact semisimple Lie group are investigated. In the process, a local variant of the Hausdorff-Young inequality on such groups is given.

INTRODUCTION

Let $1 \leq p \leq 2$. An operator space E is said to have Fourier type p with respect to the compact group G if the vector-valued Fourier transform extends to a completely bounded map

$$\mathcal{F}_{G,E} : L_E^p(G) \longrightarrow \mathcal{L}_E^{p'}(\widehat{G}),$$

where $p' = p/(p-1)$ is the exponent conjugate to p . That is, a vector-valued Hausdorff-Young inequality of exponent p is satisfied. Similarly, if we replace the operator $\mathcal{F}_{G,E}$ by its inverse, we get the notion of Fourier cotype p' of E with respect to G . Following the notation of [8], we define the constants

$$\mathcal{C}_p^1(E, G) = \|\mathcal{F}_{G,E}\|_{cb(L_E^p(G), \mathcal{L}_E^{p'}(\widehat{G}))} \quad \text{and} \quad \mathcal{C}_{p'}^2(E, G) = \|\mathcal{F}_{G,E}^{-1}\|_{cb(\mathcal{L}_E^p(\widehat{G}), L_E^{p'}(G))}.$$

The Fourier type and cotype become stronger conditions on the pair (E, G) as the exponents p and p' approach 2. This gives rise to the notions of sharp Fourier type and cotype exponents. The present paper grew out of the project to investigate the sharp Fourier type and cotype of Lebesgue spaces L^p and Schatten classes S^p , and it is a natural continuation of [8]. However, as we shall see below, some other results have appeared in the process which are interesting in their own right.

In Section 1 we recall that the natural candidates for the sharp Fourier type and cotype of L^p and S^p (where now $1 \leq p \leq \infty$) are $\min(p, p')$ and $\max(p, p')$ respectively. To show that this guess is right, one would have to show that for $1 \leq p < q \leq 2$,

$$\begin{aligned} (a) \quad \mathcal{C}_q^1(L^p(\Omega), G) &= \mathcal{C}_{q'}^2(L^{p'}(\Omega), G) = \infty, \\ (b) \quad \mathcal{C}_q^1(L^{p'}(\Omega), G) &= \mathcal{C}_{q'}^2(L^p(\Omega), G) = \infty, \end{aligned}$$

with the obvious modifications for the Schatten classes. In Section 2 we make some remarks about (a) and (b). First we show that, to have any chance of getting

Received by the editors March 22, 2002.

2000 *Mathematics Subject Classification*. Primary 43A77; Secondary 22E46, 46L07.

Key words and phrases. Sharp Fourier type and cotype, Fourier transform, operator space, compact semisimple Lie group, central function, local Hausdorff-Young inequality.

Research supported in part by the European Commission via the TMR Network "Harmonic Analysis" and by Project BFM 2001/0189, Spain.

positive answers to these questions, we have to require that the group G not be finite and the operator spaces L^p and S^p be infinite-dimensional. Then, under such assumptions, one can easily get the following inequality:

$$\mathcal{C}_q^1(L^p(\Omega), G) \geq \limsup_{n \rightarrow \infty} \mathcal{C}_q^1(l^p(n), G)$$

and the analog for $L^{p'}(\Omega)$. Therefore, the growth of $\mathcal{C}_q^1(l^p(n), G)$ and $\mathcal{C}_q^1(l^{p'}(n), G)$ provides a possible way to obtain (a) and (b). In the last part of Section 2 we analyze the vector-valued Lebesgue spaces and Schatten classes.

The growth of $\mathcal{C}_q^1(l^p(n), G)$ is investigated in Section 3. To be precise, if G stands for a compact semisimple Lie group and $1 \leq p < q \leq 2$, then there exists a constant $0 \leq \mathcal{K}(G, q) \leq 1$ such that $\mathcal{C}_q^1(l^p(n), G) \geq \mathcal{K}(G, q)n^{1/p-1/q}$ for all $n \geq 1$. If one is able to show that $\mathcal{K}(G, q) > 0$, this result gives (a). Moreover, we would obtain optimal growth, since $\mathcal{C}_q^1(l^p(n), G) \leq n^{1/p-1/q}$ for any compact group. We shall see that

$$\mathcal{K}(G, q) = \inf_{n \geq 1} \sup \left\{ \frac{\|\widehat{f}\|_{L^{q'}(\widehat{G})}}{\|f\|_{L^q(G)}} : f \text{ central, } f \in L^q(G), \text{ supp}(f) \subset \mathcal{U}_n \right\}$$

where $\{\mathcal{U}_n : n \geq 1\}$ is a basis of neighborhoods of $\mathbf{1}$, the identity element of G . The Hausdorff-Young inequality on compact groups yields $\mathcal{K}(G, q) \leq 1$. The interesting point lies in the inequality $\mathcal{K}(G, q) > 0$, which constitutes a local variant of the Hausdorff-Young inequality on G with exponent q .

Sections 4 and 5 are completely devoted to the proof of this local inequality. In the abelian setting, the particular case $G = \mathbb{T}$ was explored by Andersson in [1]. The basic idea is to consider a function $f : \mathbb{T} \rightarrow \mathbb{C}$ as a complex-valued function on \mathbb{R} supported in $[-1/2, 1/2)$. Then, by expressing the norm of \widehat{f} on $L^{q'}(\mathbb{R})$ as a Riemann sum, one obtains

$$\frac{\|\widehat{f}\|_{L^{q'}(\mathbb{R})}}{\|f\|_{L^q(\mathbb{R})}} = \lim_{k \rightarrow \infty} \frac{\|\widehat{\varphi}_k\|_{L^{q'}(\mathbb{T})}}{\|\varphi_k\|_{L^q(\mathbb{T})}}$$

where $\varphi_k(t) = k^{1/q}f(kt)$. This gives $\mathcal{K}(\mathbb{T}, q) \geq \mathcal{B}_q$, where $\mathcal{B}_q = \sqrt{q^{1/q}/q'^{1/q'}}$ stands for the Babenko-Beckner constant, see [2] and [3], but in fact the equality holds, as was proved by Sjölin in [15]. We show here that Andersson's argument, suitably modified, is also valid in the context of compact semisimple Lie groups. In Section 4 we summarize the main results of the structure and representation theory of compact semisimple Lie groups that will be used in the process. Then we use these algebraic results to get an expression for the Fourier transform of central functions $f : G \rightarrow \mathbb{C}$ in terms of the Fourier transform $\mathcal{F}_{\mathbf{T}}$ on the maximal torus \mathbf{T} of G . This will allow us to work over the maximal torus where we know that Andersson obtained a satisfactory result. However, in the non-commutative setting, the degree d_π of an irreducible representation π does not have to be 1. We shall see that this becomes a further obstacle, to be treated in Section 5. There we combine some results, such as the Weyl dimension formula, concerning the representation theory of compact semisimple Lie groups with classical harmonic analysis to avoid this difficulty.

On the other hand, if we notice that $\mathcal{C}_q^1(l^{p'}(n), G) = \mathcal{C}_q^2(l^p(n), G)$, we can understand the growth of this constant as the dual problem of the growth of $\mathcal{C}_q^1(l^p(n), G)$ in the sense that we replace the Fourier transform operator $\mathcal{F}_{G, l^p(n)}$ by its inverse.

Therefore, since the dual object is no longer a group (as it is when G is abelian) we do not have a Fourier inversion theorem and we should not expect to reconstruct the proof given in Sections 3, 4 and 5 step by step. At the time of this writing, we are not able to solve this problem, and so we pose it as follows:

Problem. Let G be any compact semisimple Lie group and let $1 \leq p < q \leq 2$. Does the estimate $\mathcal{C}_q^1(l^{p'}(n), G) \geq \mathcal{K}(G, q)n^{1/p-1/q}$ hold for some positive constant $\mathcal{K}(G, q)$ depending only on G and q ?

Finally, we point to a non-commutative notion of Rademacher type for operator spaces, see [9]. We think this notion could be helpful in order to study the growth of $\mathcal{C}_q^1(l^{p'}(n), G)$.

1. STATEMENT OF THE PROBLEM

All throughout this paper some basic notions of operator space theory and non-commutative vector-valued discrete L^p spaces will be assumed. The definitions and results about operator spaces that we are using can be found in the book of Effros and Ruan [5], while for the study of our non-commutative L^p spaces the reader is referred to [12], where Pisier analyzes them in detail. In any case, all the analytic preliminaries of this paper are summarized in [8], where we study the Fourier type and cotype of an operator space with respect to a compact group. In order to state the problem we want to solve, we begin by recalling the definitions and the main properties of Fourier type and cotype.

Let G be a compact topological group endowed with its Haar measure μ normalized so that $\mu(G) = 1$, and let $\pi \in \widehat{G}$ be an irreducible unitary representation of G of degree d_π . Here the symbol \widehat{G} stands for the dual object of G . Given an operator space E , it was shown in [8] that, by fixing a basis on the representation space of each $\pi \in \widehat{G}$, the Fourier transform operator $\mathcal{F}_{G,E}$ for functions defined on G and with values on E has the form $f \in L_E^1(G) \mapsto (\widehat{f}(\pi))_{\pi \in \widehat{G}} \in \mathcal{M}_E(\widehat{G})$, where

$$\widehat{f}(\pi) = \int_G f(g)\pi(g)^* d\mu(g) \quad \text{and} \quad \mathcal{M}_E(\widehat{G}) = \prod_{\pi \in \widehat{G}} M_{d_\pi} \otimes E.$$

Here M_n denotes the space of $n \times n$ complex matrices. Let $1 \leq p < \infty$. If $S_n^p(E)$ stands for the vector-valued Schatten class on $M_n \otimes E$, we define the spaces

$$\begin{aligned} \bullet \quad \mathcal{L}_E^p(\widehat{G}) &= \left\{ A \in \mathcal{M}_E(\widehat{G}) : \|A\|_{\mathcal{L}_E^p(\widehat{G})} = \left(\sum_{\pi \in \widehat{G}} d_\pi \|A^\pi\|_{S_{d_\pi}^p(E)}^p \right)^{1/p} < \infty \right\}, \\ \bullet \quad \mathcal{L}_E^\infty(\widehat{G}) &= \left\{ A \in \mathcal{M}_E(\widehat{G}) : \|A\|_{\mathcal{L}_E^\infty(\widehat{G})} = \sup_{\pi \in \widehat{G}} \|A^\pi\|_{S_{d_\pi}^\infty(E)} < \infty \right\}. \end{aligned}$$

We write $\mathcal{L}^p(\widehat{G})$ for the case $E = \mathbb{C}$. Finally, let $1 \leq p \leq 2$. Then by the Hausdorff-Young inequality on compact groups (see [8] or Kunze's paper [10]) it is not difficult to check that $\mathcal{F}_{G,E}(L^p(G) \otimes E) \subset \mathcal{L}^{p'}(\widehat{G}) \otimes E$ and $\mathcal{F}_{G,E}^{-1}(\mathcal{L}^p(\widehat{G}) \otimes E) \subset L^{p'}(G) \otimes E$. This motivates the following definitions.

Definition 1.1. Let $1 \leq p \leq 2$ and let p' denote its conjugate exponent. The operator space E has *Fourier type* p with respect to the compact group G if the Fourier transform operator

$$\mathcal{F}_{G,E} : L^p(G) \otimes E \rightarrow \mathcal{L}^{p'}(\widehat{G}) \otimes E$$

can be extended to a completely bounded operator from $L_E^p(G)$ into $\mathcal{L}_E^{p'}(\widehat{G})$. In that case, $\mathcal{C}_p^1(E, G)$ will stand for its *cb* norm.

Definition 1.2. In the same fashion, the operator space E has *Fourier cotype* p' with respect to the compact group G if the inverse

$$\mathcal{F}_{G,E}^{-1} : \mathcal{L}^p(\widehat{G}) \otimes E \rightarrow L^{p'}(G) \otimes E$$

can be extended to a completely bounded operator from $\mathcal{L}_E^p(\widehat{G})$ to $L_E^{p'}(G)$. As before, we shall denote its *cb* norm by $\mathcal{C}_{p'}^2(E, G)$.

One of the properties proved in [8] is that every operator space has Fourier type 1 and Fourier cotype ∞ with respect to any compact group. In particular, the complex interpolation method for operator spaces (see Pisier's work [11]) provides the following result.

Lemma 1.3. *Let $1 \leq p_1 \leq p_2 \leq 2$ and assume that E has Fourier type p_2 with respect to G . Then E has Fourier type p_1 with respect to G . Similarly, Fourier cotype p'_2 of E with respect to G implies Fourier cotype p'_1 of E with respect to G .*

Therefore, the Fourier type and cotype become stronger conditions on the pair (E, G) as the exponent p (and consequently its conjugate p') tends to 2. So Lemma 1.3 gives rise to the following definition.

Definition 1.4. The *sharp Fourier type and cotype exponents* of an operator space E with respect to the compact group G are defined respectively by

$$\begin{aligned} p_1(E, G) &= \sup\{p \leq 2 : E \text{ has Fourier type } p \text{ with respect to } G\}, \\ p_2(E, G) &= \inf\{p' \geq 2 : E \text{ has Fourier cotype } p' \text{ with respect to } G\}. \end{aligned}$$

If E has Fourier type $p_1(E, G)$ with respect to G , we say that E has *sharp Fourier type* $p_1(E, G)$. The *sharp Fourier cotype* of E is defined analogously.

In order to simplify the statement of the problem we shall need the following lemma (see [8]) which analyzes the Fourier type and cotype of the dual E^* of an operator space E with respect to a compact group G .

Lemma 1.5. *Let $1 \leq p \leq 2$ and let p' be the conjugate exponent of p . Then we have the equalities $\mathcal{C}_p^1(E^*, G) = \mathcal{C}_{p'}^2(E, G)$ and $\mathcal{C}_{p'}^2(E^*, G) = \mathcal{C}_p^1(E, G)$.*

The problem we want to investigate in this paper is how to find the sharp Fourier type and cotype of Lebesgue spaces and Schatten classes. Concerning these topics, we present here a result given in [8] from which we start out. In what follows $(\Omega, \mathcal{A}, \nu)$ will denote a σ -finite or regular measure space and $S_{\mathbb{N}}^p$ the classical Schatten class over the space of compact operators on l^2 .

Theorem 1.6. *If $1 \leq p \leq \infty$, then the spaces $L^p(\Omega)$, $S_{\mathbb{N}}^p$ and $S_{\mathbb{N}}^p$ have Fourier type $\min(p, p')$ and Fourier cotype $\max(p, p')$. In fact, the vector-valued Fourier transform, or its inverse, is a complete contraction in each of the cases considered.*

Therefore, if we consider two exponents p and q such that $1 \leq p < q \leq 2$, we would like to find conditions on G and Ω under which

$$\begin{aligned} (a) \quad \mathcal{C}_q^1(L^p(\Omega), G) &= \mathcal{C}_{q'}^2(L^{p'}(\Omega), G) = \infty, \\ (b) \quad \mathcal{C}_q^1(L^{p'}(\Omega), G) &= \mathcal{C}_{q'}^2(L^p(\Omega), G) = \infty, \end{aligned}$$

with the obvious modifications for the Schatten classes.

2. SOME REMARKS ABOUT THE PROBLEM

In this section we shall make some remarks about the problem we have just stated. We begin by showing some necessary conditions that should hold to obtain a positive answer to our question. Second, we wonder about sufficient conditions that we shall work with in the rest of this paper. Finally, we study what happens if we consider vector-valued L^p spaces (or Bochner-Lebesgue spaces) and vector-valued Schatten classes.

2.1. Necessary conditions. The first necessary condition that we shall present is on the compact group G . We have to exclude **finite groups** from our treatment, since, as we shall see immediately, every operator space E has sharp Fourier type and cotype 2 with respect to any finite group. Anyway, the next result is a bit more accurate.

Proposition 2.1. *For a finite group G , every operator space E satisfies the estimates $\mathcal{C}_p^1(E, G), \mathcal{C}_{p'}^2(E, G) \leq |G|^{1/p'}$ for $1 \leq p \leq 2$.*

Proof. Let us assume that $\mathcal{C}_2^1(E, G) \leq |G|^{1/2}$ for every operator space E ; then we have $\mathcal{C}_2^2(E, G) = \mathcal{C}_2^1(E^*, G) \leq |G|^{1/2}$ by duality. The desired estimates are then obtained by complex interpolation from the equalities $\mathcal{C}_1^1(E, G) = \mathcal{C}_\infty^2(E, G) = 1$ (proved in [8]) and the case $p = 2$. Therefore we focus our attention on the case $p = 2$. It suffices to check that for all $m \geq 1$ and any family of functions $\{f_{ij} : G \rightarrow E\}_{1 \leq i, j \leq m}$,

$$\left(\sum_{\pi \in \hat{G}} d_\pi \left\| \begin{pmatrix} \widehat{f_{ij}}(\pi) \end{pmatrix} \right\|_{S_{d_\pi m}^2(E)}^2 \right)^{1/2} \leq |G|^{1/2} \left\| \begin{pmatrix} f_{ij} \end{pmatrix} \right\|_{S_m^2(L_E^2(G))}.$$

But if $G = \{g_1, g_2, \dots, g_n\}$, then

$$\begin{aligned} \left\| \begin{pmatrix} \widehat{f_{ij}}(\pi) \end{pmatrix} \right\|_{S_{d_\pi m}^2(E)} &= \left\| \begin{pmatrix} \frac{1}{n} \sum_{k=1}^n f_{ij}(g_k) \pi(g_k)^* \end{pmatrix} \right\|_{S_{d_\pi m}^2(E)} \\ &\leq \frac{1}{n} \sum_{k=1}^n \|\pi(g_k)^*\|_{S_{d_\pi}^2} \left\| \begin{pmatrix} f_{ij}(g_k) \end{pmatrix} \right\|_{S_m^2(E)} \\ &\leq d_\pi^{1/2} \left\| \begin{pmatrix} f_{ij} \end{pmatrix} \right\|_{S_m^2(L_E^2(G))}. \end{aligned}$$

Therefore we obtain

$$\left\| \begin{pmatrix} \widehat{f_{ij}} \end{pmatrix} \right\|_{S_m^2(L_E^2(\hat{G}))} \leq \sqrt{\sum_{\pi \in \hat{G}} d_\pi^2} \left\| \begin{pmatrix} f_{ij} \end{pmatrix} \right\|_{S_m^2(L_E^2(G))},$$

and, since $\sum_{\pi \in \hat{G}} d_\pi^2 = |G|$ by the Peter-Weyl theorem, we are done. \square

Next we show that we cannot work with measure spaces $(\Omega, \mathcal{A}, \nu)$ that are a **union of finitely many ν -atoms**. Before that we need to define the *cb* distance between two operator spaces. It is due to Pisier, and it constitutes the analog of the Banach-Mazur distance between two Banach spaces in the context of operator space theory. Given two operator spaces E_1 and E_2 , we define their *cb* distance by the relation $d_{cb}(E_1, E_2) = \inf\{\|u\|_{cb(E_1, E_2)} \|u^{-1}\|_{cb(E_2, E_1)}\}$ where the infimum runs over all complete isomorphisms $u : E_1 \rightarrow E_2$. The following result (also extracted

from [8]) relates the Fourier type and cotype of two operator spaces E_1 and E_2 to their cb distance.

Lemma 2.2. *Let $1 \leq p \leq 2$ and let E_1, E_2 be operator spaces. Then*

$$\mathcal{C}_p^1(E_2, G) \leq d_{cb}(E_1, E_2) \mathcal{C}_p^1(E_1, G)$$

and

$$\mathcal{C}_{p'}^2(E_2, G) \leq d_{cb}(E_1, E_2) \mathcal{C}_{p'}^2(E_1, G).$$

Proposition 2.3. *Let $1 \leq p < q \leq 2$ and assume that $(\Omega, \mathcal{A}, \nu)$ is a union of finitely many ν -atoms. Then every compact group G satisfies the following estimates:*

$$\begin{aligned} \mathcal{C}_q^1(L^p(\Omega), G) &= \mathcal{C}_{q'}^2(L^{p'}(\Omega), G) \leq \nu(\Omega)^{1/p-1/q}, \\ \mathcal{C}_q^1(L^{p'}(\Omega), G) &= \mathcal{C}_{q'}^2(L^p(\Omega), G) \leq \nu(\Omega)^{1/q'-1/p'}. \end{aligned}$$

In fact, since $1/p - 1/q = 1/q' - 1/p'$, we have the same bound for both cb norms.

Proof. Applying Lemma 2.2 and the last part of Theorem 1.6, we get the estimates $\mathcal{C}_q^1(L^p(\Omega), G) \leq d_{cb}(L^p(\Omega), L^q(\Omega))$ and $\mathcal{C}_q^1(L^{p'}(\Omega), G) \leq d_{cb}(L^{p'}(\Omega), L^{q'}(\Omega))$. On the other hand, it is straightforward to check that, for $1 \leq p_1 < p_2 \leq \infty$ and such a measure space $(\Omega, \mathcal{A}, \nu)$, we have $d_{cb}(L^{p_1}(\Omega), L^{p_2}(\Omega)) \leq \nu(\Omega)^{1/p_1-1/p_2}$. \square

In other words, we do not allow **finite-dimensional** Lebesgue spaces. Since the cb distance between two Schatten classes of the same finite dimension is also finite, the arguments used in Proposition 2.3, Theorem 1.6, and Lemma 2.2 are also valid to show that the only Schatten classes that could make Theorem 1.6 sharp are those of infinite dimension.

2.2. Sufficient conditions. The Fourier type and cotype of the subspaces of a given operator space E are bounded above by the respective type and cotype of E . The proof of this result is straightforward, see [8].

Lemma 2.4. *Let $1 \leq p \leq 2$ and let F be a closed subspace of E . Then we have the estimates $\mathcal{C}_p^1(F, G) \leq \mathcal{C}_p^1(E, G)$ and $\mathcal{C}_{p'}^2(F, G) \leq \mathcal{C}_{p'}^2(E, G)$.*

After the conditions above, we shall work in the sequel with infinite compact groups and infinite-dimensional Lebesgue spaces and Schatten classes. Since the measure space $(\Omega, \mathcal{A}, \nu)$ is no longer a union of finitely many ν -atoms, we obtain that the n -dimensional space $l^p(n)$ is a closed subspace of $L^p(\Omega)$ for all $n \geq 1$ and any $1 \leq p \leq \infty$. Moreover, recalling that the subspace of diagonal matrices of S_n^p is completely isomorphic to $l^p(n)$, we deduce that the same happens for the Schatten classes S_N^p . Hence sharpness of Theorem 1.6 will be guaranteed if, for $1 \leq p < q \leq 2$, we have

$$\begin{aligned} (a') \quad \mathcal{C}_q^1(l^p(n), G) &= \mathcal{C}_{q'}^2(l^{p'}(n), G) \longrightarrow \infty && \text{as } n \rightarrow \infty, \\ (b') \quad \mathcal{C}_q^1(l^{p'}(n), G) &= \mathcal{C}_{q'}^2(l^p(n), G) \longrightarrow \infty && \text{as } n \rightarrow \infty. \end{aligned}$$

Therefore our aim from now on will be the study of the growth of the constants $\mathcal{C}_q^1(l^p(n), G)$ and $\mathcal{C}_q^1(l^{p'}(n), G)$. The first remark about these constants that we can already make is that both have a common upper bound:

$$\mathcal{C}_q^1(l^p(n), G), \mathcal{C}_q^1(l^{p'}(n), G) \leq n^{1/p-1/q}.$$

This is an obvious consequence of Proposition 2.3.

2.3. Vector-valued spaces. Theorem 1.6 was also studied in [8] for vector-valued spaces. Here is the statement of the result obtained there.

Theorem 2.5. *Let $1 \leq p \leq \infty$ and let E be an operator space having Fourier type $\min(p, p')$ (respectively, Fourier cotype $\max(p, p')$) with respect to G . Then the spaces $L_E^p(\Omega)$, $S_n^p(E)$ and $S_{\mathbb{N}}^p(E)$ have Fourier type $\min(p, p')$ (respectively, Fourier cotype $\max(p, p')$) with respect to G .*

Let $1 \leq p \leq \infty$ and $\min(p, p') < q \leq 2$. Let E be as in Theorem 2.5. Then Lemma 2.4 gives the following estimates:

$$\begin{aligned} \mathcal{C}_q^1(L_E^p(\Omega), G) &\geq \mathcal{C}_q^1(L^p(\Omega), G), & \mathcal{C}_q^1(L_E^p(\Omega), G) &\geq \mathcal{C}_q^1(E, G), \\ \mathcal{C}_{q'}^2(L_E^p(\Omega), G) &\geq \mathcal{C}_{q'}^2(L^p(\Omega), G), & \mathcal{C}_{q'}^2(L_E^p(\Omega), G) &\geq \mathcal{C}_{q'}^2(E, G), \end{aligned}$$

with the obvious modifications for the Schatten classes. Hence we have shown that sharp Fourier type or cotype of $L^p(\Omega)$ (respectively $S_{\mathbb{N}}^p$) provides sharp Fourier type or cotype of $L_E^p(\Omega)$ (respectively $S_{\mathbb{N}}^p(E)$) with respect to G . Also, the same conclusion is obtained assuming sharp Fourier type or cotype of E . In particular, the sufficient condition given above also works for vector-valued spaces. Therefore we focus our attention on the growth of the constants $\mathcal{C}_q^1(l^p(n), G)$ and $\mathcal{C}_q^1(l^{p'}(n), G)$.

3. ON THE GROWTH OF $\mathcal{C}_q^1(l^p(n), G)$.

We shall assume in what follows that G is a **compact semisimple Lie group**. Semisimplicity is an essential assumption in the arguments that we shall be using. Anyway, for the moment, the only property of such groups that we shall apply is the existence of a maximal torus \mathbf{T} in G . The following result gives, in particular, statement (a) in Section 1, with the obvious modifications for Schatten classes, whenever we work with infinite-dimensional operator spaces and compact semisimple Lie groups.

Theorem 3.1. *Let $1 \leq p < q \leq 2$ and let G be a compact semisimple Lie group. Then there exists a constant $0 < \mathcal{K}(G, q) \leq 1$, depending on G and q , such that for all $n \geq 1$,*

$$\mathcal{K}(G, q)n^{1/p-1/q} \leq \mathcal{C}_q^1(l^p(n), G) \leq n^{1/p-1/q}.$$

In particular, we observe that the growth of $\mathcal{C}_q^1(l^p(n), G)$ is optimal for compact semisimple Lie groups. The proof of this result starts by applying the existence of a maximal torus \mathbf{T} to consider a countable family $\{g_k : k \geq 1\}$ of pairwise commuting elements of G ; just take $g_k \in \mathbf{T}$. For every $n \geq 1$ we take \mathcal{U}_n to be a neighborhood of $\mathbf{1}$ (the identity element of G) satisfying

$$g_j^{-1}\mathcal{U}_n \cap g_k^{-1}\mathcal{U}_n = \emptyset \quad \text{for } 1 \leq j, k \leq n \text{ and } j \neq k.$$

We recall here that we can always consider a central function f_n supported in \mathcal{U}_n and belonging to $L^q(G)$. For example, take \mathcal{U}_n to be invariant under conjugations (see Lemma (5.24) of [6]) and $f_n = 1_{\mathcal{U}_n}$ where $1_{\mathcal{U}}$ stands for the characteristic function of \mathcal{U} . Henceforth f_n will be a central function in $L^q(G)$ supported in \mathcal{U}_n , to be fixed later. Then we define the function $\Phi_n : G \rightarrow \mathbb{C}^n$ by $\Phi_n(g) = (f_n(g_1g), f_n(g_2g), \dots, f_n(g_ng))$. We obviously have the estimate

$$\mathcal{C}_q^1(l^p(n), G) \geq \frac{\|\widehat{\Phi}_n\|_{\mathcal{L}_{l^p(n)}^{q'}(\widehat{G})}}{\|\Phi_n\|_{L_{l^p(n)}^q(G)}}.$$

So it suffices to prove that this quotient is bounded below by $\mathcal{K}(G, q)n^{1/p-1/q}$. The following lemma will be very helpful for that purpose.

Lemma 3.2. *Let $1 \leq p_1, p_2 \leq \infty$, $\pi \in \widehat{G}$ and $n \geq 1$. Consider the matrix-valued vector $A_{\pi,n} = (\pi(g_1), \pi(g_2), \dots, \pi(g_n))$. Then*

$$\|A_{\pi,n}\|_{l^{p_2}_{S^{p_1}_{d_\pi}}(n)} = \|A_{\pi,n}\|_{S^{p_2}_{d_\pi}(l^{p_1}(n))} = n^{1/p_1} d_\pi^{1/p_2}.$$

Proof. Since g_1, g_2, \dots, g_n are pairwise commuting, there exists a basis of \mathbb{C}^{d_π} of common eigenvectors of $\pi(g_1), \pi(g_2), \dots, \pi(g_n)$. Therefore, in that basis, all these matrices are diagonal:

$$\pi(g_k) = \begin{pmatrix} \theta_1^k & & \\ & \ddots & \\ & & \theta_{d_\pi}^k \end{pmatrix}.$$

Moreover, $|\theta_j^k| = 1$ for $1 \leq j \leq d_\pi$ because of the unitarity of $\pi(g_k)$. Hence, applying the complete isometry (see Corollary (1.3) of [12]) between $l^{p_2}_E(d_\pi)$ and the subspace of diagonal matrices of $S^{p_2}_{d_\pi}(E)$, we easily obtain the desired equality. \square

(i) **The value of $\|\widehat{\Phi}_n\|_{\mathcal{L}^{q'}_{l^{q'}(n)}(\widehat{G})}$.** We begin by recalling that, since f_n is central,

$$\widehat{f}_n(\pi) = \frac{1}{d_\pi} \int_G f_n(g) \overline{\chi_\pi(g)} d\mu(g) 1_{d_\pi} = \gamma_{\pi,n} 1_{d_\pi}$$

by Schur’s lemma. Here χ_π is the irreducible character associated to π and 1_m denotes the identity matrix of order $m \times m$. On the other hand, $f_n(g_k \cdot)$ is the translation by g_k of f_n ; therefore,

$$\widehat{\Phi}_n(\pi) = \frac{1}{d_\pi} \int_G f_n(g) \overline{\chi_\pi(g)} d\mu(g) \ (\pi(g_1), \pi(g_2), \dots, \pi(g_n)) = \gamma_{\pi,n} A_{\pi,n}.$$

So we get, by Lemma 3.2, the following equality:

$$\begin{aligned} \|\widehat{\Phi}_n\|_{\mathcal{L}^{q'}_{l^{q'}(n)}(\widehat{G})} &= \left(\sum_{\pi \in \widehat{G}} d_\pi |\gamma_{\pi,n}|^{q'} \|A_{\pi,n}\|_{S^{q'}_{d_\pi}(l^{q'}(n))}^{q'} \right)^{1/q'} \\ &= n^{1/p} \left(\sum_{\pi \in \widehat{G}} d_\pi^2 |\gamma_{\pi,n}|^{q'} \right)^{1/q'} = n^{1/p} \|\widehat{f}_n\|_{\mathcal{L}^{q'}(\widehat{G})}. \end{aligned}$$

(ii) **The value of $\|\Phi_n\|_{L^q_{l^p(n)}(G)}$.** We have

$$\begin{aligned} \|\Phi_n\|_{L^q_{l^p(n)}(G)} &= \left(\int_G \left(\sum_{k=1}^n |f_n(g_k g)|^p \right)^{q/p} d\mu(g) \right)^{1/q} \\ &= \left(\sum_{k=1}^n \|f_n(g_k \cdot)\|_{L^q(G)}^q \right)^{1/q} = n^{1/q} \|f_n\|_{L^q(G)} \end{aligned}$$

since the sets $\{g_k^{-1} \mathcal{U}_n : 1 \leq k \leq n\}$ are pairwise disjoint.

In summary, we have obtained that $C^1_q(l^p(n), G) \geq \mathcal{K}(G, q, n)n^{1/p-1/q}$ where the constant $\mathcal{K}(G, q, n)$ is given by

$$\mathcal{K}(G, q, n) = \frac{\|\widehat{f}_n\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|f_n\|_{L^q(G)}}.$$

If we define $\mathcal{K}(G, q) = \inf_{n \geq 1} \mathcal{K}(G, q, n)$, it is obvious that $\mathcal{K}(G, q) \leq 1$ by the Hausdorff-Young inequality on compact groups. Thus it remains to check that $\mathcal{K}(G, q) > 0$. For that aim, since we have not fixed f_n yet, we need to see that

$$\inf_{n \geq 1} \sup \left\{ \frac{\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|f\|_{L^q(G)}} : f \text{ central, } f \in L^q(G), \text{ supp}(f) \subset \mathcal{U}_n \right\} > 0.$$

We shall prove this fact in Section 5 where we study the supremum of the Hausdorff-Young quotient for central functions supported in arbitrary small sets. As we shall see immediately, semisimplicity of G will be essential in our proof.

4. A SIMPLE EXPRESSION FOR THE FOURIER TRANSFORM OF CENTRAL FUNCTIONS

In this section we apply some basic results concerning the structure and representation theory of compact semisimple Lie groups to provide a simple expression for the Fourier transform of central functions defined on such groups. These algebraic preliminaries can be found in Simon's book [14] or alternatively in [7], but we summarize the main topics here. Let G be a compact semisimple Lie group and let \mathfrak{g} be its Lie algebra. In what follows we choose once and for all an explicit maximal torus \mathbf{T} in G , while \mathfrak{h} will stand for its Lie algebra. That is, \mathfrak{h} is the Cartan subalgebra of \mathfrak{g} . The rank of G will be denoted by r ; in particular, $\mathbf{T} \simeq \mathbb{T}^r$ where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with its natural group structure. Also, as is customary, we consider the complexification $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \oplus i\mathfrak{g}$ (with complex conjugates taken so that $\mathfrak{g}_{\mathbb{R}} = \{Z \in \mathfrak{g}_{\mathbb{C}} : Z = \overline{Z}\} = i\mathfrak{g}$ and similarly $\mathfrak{h}_{\mathbb{R}} = i\mathfrak{h}$) with the complex inner product $\langle \cdot, \cdot \rangle$ induced by the Killing form. We also recall that the Weyl group \mathcal{W}_G associated to G can be seen as a set of $r \times r$ unitary matrices W (isometries on $\mathfrak{h}_{\mathbb{R}}$) with integer entries and $\det W = \pm 1$. In particular, the set $\mathcal{W}_G^* = \{W^t : W \in \mathcal{W}_G\}$ becomes a set of isometries on $\mathfrak{h}_{\mathbb{R}}^*$. The symbol \mathcal{R} will stand for the set of roots, and, if we take $H_0 \in \mathfrak{h}_{\mathbb{R}}$ such that $\alpha(H_0) \neq 0$ for any root α , the symbol $\mathcal{R}^+ = \{\alpha \in \mathcal{R} : \alpha(H_0) > 0\}$ denotes the set of positive roots. Finally, we shall write $\Lambda_{\mathbf{W}}$ and Λ_{DW} for the weight lattice and the set of dominant weights, respectively.

Now that we have fixed some notation, let us consider a central function $f : G \rightarrow \mathbb{C}$ and a dominant weight $\lambda \in \Lambda_{\text{DW}}$. By the dominant weight theorem there exists a unique $\pi_{\lambda} \in \widehat{G}$ associated to λ and, since f is central, we can write, by Schur's lemma,

$$\widehat{f}(\pi_{\lambda}) = \frac{1}{d_{\lambda}} \int_G f(g) \overline{\chi_{\lambda}(g)} d\mu(g) 1_{d_{\lambda}}$$

where d_{λ} is the degree of π_{λ} , χ_{λ} is the character of π_{λ} and 1_m denotes the $m \times m$ identity matrix. We now recall the definition of the functions A_{β} appearing in the Weyl character formula. Given $\beta \in \mathfrak{h}_{\mathbb{R}}^*$, we define the functions $\exp_{\beta} : \mathfrak{h}_{\mathbb{R}} \rightarrow \mathbb{C}$ and $A_{\beta} : \mathfrak{h}_{\mathbb{R}} \rightarrow \mathbb{C}$ by the relations

$$\begin{aligned} \exp_{\beta}(H) &= e^{2\pi i \langle \beta, H \rangle}, \\ A_{\beta}(H) &= \sum_{W \in \mathcal{W}_G} \det W \exp_{\beta}(W(H)). \end{aligned}$$

The maximal torus \mathbf{T} is isomorphic via the exponential mapping to the quotient space $\mathfrak{h}_{\mathbb{R}}/L_{\mathbf{W}}$, where $L_{\mathbf{W}}$ is the set of those $H \in \mathfrak{h}_{\mathbb{R}}$ satisfying $\exp(2\pi i H) = 1$.

That is, L_W is the dual lattice of Λ_W . Therefore, the functions \exp_β and A_β are well-defined functions on \mathbf{T} if and only if $\beta \in \Lambda_W$. As we know, the integral form

$$\delta = \frac{1}{2} \sum_{\alpha \in \mathcal{R}^+} \alpha$$

is not necessarily a weight, and so the functions \exp_δ and A_δ could be not well-defined on \mathbf{T} . To avoid this difficulty we assume for the moment that G is **simply connected**; this condition on G assures that $\delta \in \Lambda_W$. Hence, applying consecutively the Weyl integration formula and the Weyl character formula, we get

$$\begin{aligned} \widehat{f}(\pi_\lambda) &= \frac{1}{d_\lambda |\mathcal{W}_G|} \int_{\mathbf{T}} f(t) \overline{\chi_\lambda(t)} |A_\delta(t)|^2 dm(t) 1_{d_\lambda} \\ &= \frac{1}{d_\lambda |\mathcal{W}_G|} \int_{\mathbf{T}} f(t) A_\delta(t) \overline{A_{\lambda+\delta}(t)} dm(t) 1_{d_\lambda} \end{aligned}$$

where m denotes the Haar measure on \mathbf{T} normalized so that $m(\mathbf{T}) = 1$. Now, if we write $A_{\lambda+\delta}$ as a linear combination of exponentials, we obtain

$$\begin{aligned} \widehat{f}(\pi_\lambda) &= \frac{1}{d_\lambda |\mathcal{W}_G|} \sum_{W \in \mathcal{W}_G} \det W \int_{\mathbf{T}} f(t) A_\delta(t) \exp_{-(\lambda+\delta)}(W(t)) dm(t) 1_{d_\lambda} \\ &= \frac{1}{d_\lambda} \int_{\mathbf{T}} f(t) A_\delta(t) \exp_{-(\lambda+\delta)}(t) dm(t) 1_{d_\lambda}, \end{aligned}$$

since $A_\delta(W(t)) = \det W A_\delta(t)$ and $f(W(t)) = f(t)$. We recall that, taking coordinates with respect to the basis $\{\omega_1, \omega_2, \dots, \omega_r\}$ of fundamental weights, any weight $\lambda \in \Lambda_W$ has integer coordinates. Therefore, we can understand the last expression as the Fourier transform of $f A_\delta$ on the maximal torus \mathbf{T} evaluated at $\lambda + \delta$. Hence we have

$$(1) \qquad \widehat{f}(\pi_\lambda) = \frac{1}{d_\lambda} \mathcal{F}_{\mathbf{T}}(f A_\delta)(\lambda + \delta) 1_{d_\lambda},$$

for $f : G \rightarrow \mathbb{C}$ central and G any compact semisimple simply connected Lie group. When G is not simply connected, a more careful approach is needed. We have $W^t(\delta) \pm \delta \in \Lambda_W$ for all $W \in \mathcal{W}_G$. In particular, we note that

$$\exp_{\pm \delta} A_{\lambda+\delta} = \sum_{W \in \mathcal{W}_G} \det W \exp_{W^t(\lambda+\delta) \pm \delta}$$

is a well-defined function on \mathbf{T} for all $\lambda \in \Lambda_{DW}$. This remark allows us to write $\overline{\chi_\lambda} |A_\delta|^2 = (\exp_\delta \overline{A_{\lambda+\delta}})(\exp_{-\delta} A_\delta)$ as a well-defined function on \mathbf{T} . Henceforth, applying again Schur's lemma, the Weyl integration formula and the Weyl character formula, we get

$$\begin{aligned} \widehat{f}(\pi_\lambda) &= \frac{1}{d_\lambda |\mathcal{W}_G|} \sum_{W \in \mathcal{W}_G} \det W \int_{\mathbf{T}} f(t) (\exp_{-\delta} A_\delta)(t) \exp_{\delta - W^t(\lambda+\delta)}(t) dm(t) 1_{d_\lambda} \\ &= \frac{1}{d_\lambda} \int_{\mathbf{T}} f(t) (\exp_{-\delta} A_\delta)(t) \exp_{-\lambda}(t) dm(t) 1_{d_\lambda} \end{aligned}$$

where the last equality follows from the change of variable $t \mapsto W^t(t)$. That is, we have shown that

$$(2) \qquad \widehat{f}(\pi_\lambda) = \frac{1}{d_\lambda} \mathcal{F}_{\mathbf{T}}(f B_\delta)(\lambda) 1_{d_\lambda}$$

where $B_\delta = \exp_{-\delta} A_\delta$. This expression is now valid for any compact semisimple Lie group, and it coincides with (1) for simply connected ones.

5. A LOCAL VARIANT OF THE HAUSDORFF-YOUNG INEQUALITY ON COMPACT SEMISIMPLE LIE GROUPS

As we mentioned in the introduction, this section is devoted to the proof of a local variant of the Hausdorff-Young inequality on compact semisimple Lie groups. We recall that this result provides the relation $\mathcal{K}(G, q) > 0$ for $1 \leq q \leq 2$, which we needed in Section 3.

Theorem 5.1. *Let $1 \leq q \leq 2$ and let G be a compact semisimple Lie group. Then there exists a constant $0 < \mathcal{K}(G, q) \leq 1$ such that, for any open set $\mathcal{U} \subset G$, we have*

$$\sup \left\{ \frac{\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|f\|_{L^q(G)}} : f \text{ central, } f \in L^q(G), \text{ supp}(f) \subset \mathcal{U} \right\} \geq \mathcal{K}(G, q).$$

Since the norms of \widehat{f} and f on $\mathcal{L}^{q'}(\widehat{G})$ and $L^q(G)$ respectively do not change under translations of f , we can assume, without loss of generality, that \mathcal{U} is a neighborhood of $\mathbf{1}$. Before proving Theorem 5.1 we need some auxiliary results. Let us assume that G is simply connected and let $f : G \rightarrow \mathbb{C}$ be a central function. A quick look at relation (1) allows us to write

$$(3) \quad \widehat{f}(\pi_\lambda) = \frac{1}{d_\lambda} \det W \mathcal{F}_{\mathbf{T}}(f A_\delta)(W^t(\lambda + \delta)) \mathbf{1}_{d_\lambda}$$

for all $W \in \mathcal{W}_G$. On the other hand, let us denote by P_α the hyperplane of $\mathfrak{h}_{\mathbb{R}}^*$ orthogonal to α with respect to the complex inner product given by the Killing form. The infinitesimal Cartan-Stiefel diagram is then given by the expression

$$\mathbf{P} = \bigcup_{\alpha \in \mathcal{R}} P_\alpha.$$

Lemma 5.2. *Let G be a compact semisimple simply connected Lie group. Then we have $\{W^t(\lambda + \delta) : W \in \mathcal{W}_G, \lambda \in \Lambda_{\text{DW}}\} = \Lambda_{\mathbf{W}} \setminus \mathbf{P}$. Moreover, the mapping $(W, \lambda) \in \mathcal{W}_G \times \Lambda_{\text{DW}} \mapsto W^t(\lambda + \delta) \in \Lambda_{\mathbf{W}} \setminus \mathbf{P}$ is injective.*

Proof. Since G is simply connected, we have that $\{\lambda + \delta : \lambda \in \Lambda_{\text{DW}}\} = \Lambda_{\mathbf{W}} \cap \mathbf{C}^{\text{int}}$. Here \mathbf{C} stands for the fundamental Weyl chamber and \mathbf{C}^{int} for its interior. Now, since \mathbf{P} and $\Lambda_{\mathbf{W}}$ are invariant under the action of \mathcal{W}_G^* and for any Weyl chamber \mathbf{C} there exists a unique $W \in \mathcal{W}_G$ such that $W^t(\mathbf{C}) = \mathbf{C}$, we obtain the desired equality. Finally, the injectivity follows from the uniqueness mentioned above. \square

Proposition 5.3. *Let G be a compact semisimple simply connected Lie group and let $f : G \rightarrow \mathbb{C}$ be a central function. Then there exists a constant $\mathcal{A}(G, q)$, depending on G and q , such that*

$$\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})} = \mathcal{A}(G, q) \left[\sum_{\lambda \in \Lambda_{\mathbf{W}} \setminus \mathbf{P}} \frac{|\mathcal{F}_{\mathbf{T}}(f A_\delta)(\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda \rangle|^{q'-2}} \right]^{1/q'}.$$

Proof. Since f is central and G is simply connected, we can apply (3) to obtain

$$\begin{aligned}\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})} &= \left[\sum_{\lambda \in \Lambda_{\text{DW}}} d_\lambda \|\widehat{f}(\pi_\lambda)\|_{S^{q'}_{d_\lambda}}^{q'} \right]^{1/q'} \\ &= \left[\frac{1}{|\mathcal{W}_G|} \sum_{W \in \mathcal{W}_G} \sum_{\lambda \in \Lambda_{\text{DW}}} d_\lambda \left| \frac{1}{d_\lambda} \mathcal{F}_{\mathbf{T}}(fA_\delta)(W^t(\lambda + \delta)) \right|^{q'} \|1_{d_\lambda}\|_{S^{q'}_{d_\lambda}}^{q'} \right]^{1/q'}.\end{aligned}$$

Moreover, the Weyl dimension formula for d_λ gives

$$\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})} = \left[\frac{1}{|\mathcal{W}_G|} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \delta \rangle|^{q'-2} \sum_{W \in \mathcal{W}_G} \sum_{\lambda \in \Lambda_{\text{DW}}} \frac{|\mathcal{F}_{\mathbf{T}}(fA_\delta)(W^t(\lambda + \delta))|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda + \delta \rangle|^{q'-2}} \right]^{1/q'}.$$

Finally, we observe that

$$\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda + \delta \rangle| = \prod_{\alpha \in \mathcal{R}} |\langle W(\alpha), \lambda + \delta \rangle|^{1/2} = \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, W^t(\lambda + \delta) \rangle|,$$

since any $W \in \mathcal{W}_G$ is a permutation of the set of roots. Therefore, by Lemma 5.2, we have

$$\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})} = \left[\frac{1}{|\mathcal{W}_G|} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \delta \rangle|^{q'-2} \sum_{\lambda \in \Lambda_{\mathbf{W}} \setminus \mathbf{P}} \frac{|\mathcal{F}_{\mathbf{T}}(fA_\delta)(\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda \rangle|^{q'-2}} \right]^{1/q'}.$$

The proof is completed just by taking

$$A(G, q) = \left[\frac{1}{|\mathcal{W}_G|} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \delta \rangle|^{q'-2} \right]^{1/q'}.$$

□

We are now ready to give the proof of Theorem 5.1 for simply connected groups. Let $\{H_1, H_2, \dots, H_r\}$ be the predual basis of the fundamental weights; any element of $L_{\mathbf{W}}$ can be written as a linear combination of H_1, H_2, \dots, H_r with integer coefficients. Then, since $\mathbf{T} \simeq \mathfrak{h}_{\mathbb{R}}/L_{\mathbf{W}}$, we can regard \mathbf{T} as the subset of $\mathfrak{h}_{\mathbb{R}}$ given by

$$\mathfrak{T} = \left\{ \sum_{k=1}^r x_k H_k : -1/2 \leq x_k < 1/2 \right\}.$$

On the other hand, let us fix a bounded central function $f_0 : G \rightarrow \mathbb{C}$; then f_0 can be understood as a function on \mathbf{T} invariant under the action of \mathcal{W}_G . Now, since the Weyl group is generated by a set of reflections in $\mathfrak{h}_{\mathbb{R}}$, f_0 can be regarded as a complex-valued function on $\mathfrak{h}_{\mathbb{R}}$, supported in \mathfrak{T} and symmetric under such reflections. Let us recall that $\{\omega_1, \omega_2, \dots, \omega_r\}$ stands for the basis of fundamental weights. Let $\tau = 1 - 2/q'$; the way we have interpreted the function f_0 allows us to define the function

$$I_\tau(\widehat{f_0 A_\delta}) : \mathfrak{h}_{\mathbb{R}}^* \longrightarrow \mathbb{C}$$

as

$$\widehat{I_\tau(f_0 A_\delta)}(\xi) = \frac{1}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \xi \rangle|^\tau} \mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 A_\delta)(\xi) \quad \text{where} \quad \xi = \sum_{k=1}^r \xi_k \omega_k.$$

Remark 5.4. The motivation for the notation is that in a classical group such as $SU(2)$ the function just defined is nothing but the Fourier transform of the fractional integral operator

$$I_\tau(f)(x) = \frac{1}{\Gamma(\tau)} \int_{-\infty}^x f(y)(x-y)^{\tau-1} dy$$

acting on $f_0 A_\delta$. Here lies the main difference with the commutative case (where a Hausdorff-Young inequality of local type has been already investigated, see [1]) since the presence of the degrees d_λ (as a product in Proposition 5.3 by the Weyl dimension formula) requires the presence of a factor of $\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f_0 A_\delta)$. This fact does not happen in the commutative case, since $d_\lambda = 1$ for all $\lambda \in \Lambda_{\text{DW}}$.

Lemma 5.5. *Let G be a compact semisimple simply connected Lie group and let $f : G \rightarrow \mathbb{C}$ be a central function. Then $\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f A_\delta)(\xi) = 0$ for all $\xi \in \mathbf{P}$.*

Proof. If $\xi \in \mathbf{P}$, there exists a root α such that $\xi \in P_\alpha$. Let S_α be the reflection in P_α ; then $\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f A_\delta)(\xi) = \det S_\alpha \mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f A_\delta)(S_\alpha(\xi)) = -\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f A_\delta)(\xi)$, since, as we know, $S_\alpha \in \mathcal{W}_G^*$. \square

The function $\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f_0 A_\delta)$ is analytic, since $f_0 A_\delta$ has compact support and, by Lemma 5.5, it vanishes at

$$\mathbf{P} = \{\xi \in \mathfrak{h}_\mathbb{R}^* : \prod_{\alpha \in \mathcal{R}^+} \langle \alpha, \xi \rangle = 0\}.$$

In particular, since $0 \leq \tau < 1$, $\widehat{I_\tau(f_0 A_\delta)}$ is continuous and takes the value 0 on \mathbf{P} . Now we write the norm of this function in terms of a Riemann sum:

$$\|\widehat{I_\tau(f_0 A_\delta)}\|_{L^{q'}(\mathfrak{h}_\mathbb{R}^*)} = \lim_{k \rightarrow \infty} \left[\sum_{\lambda \in \Lambda_W} \frac{V_G}{k^r} \frac{|\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f_0 A_\delta)(k^{-1}\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, k^{-1}\lambda \rangle|^{\tau q'}} \right]^{1/q'},$$

where V_G denotes the volume of a cell of Λ_W . Moreover, $\phi_k(x) = k^\sigma f_0(kx) A_\delta(kx)$ is supported in \mathfrak{T} and the relation $\mathcal{F}_{\mathfrak{h}_\mathbb{R}}(f_0 A_\delta)(k^{-1}\lambda) = k^{r-\sigma} \mathcal{F}_\mathbf{T}(\phi_k)(\lambda)$ is satisfied for all $\lambda \in \Lambda_W$. Taking $\sigma = \tau|\mathcal{R}^+| + r/q$, we obtain

$$\|\widehat{I_\tau(f_0 A_\delta)}\|_{L^{q'}(\mathfrak{h}_\mathbb{R}^*)} = V_G^{1/q'} \lim_{k \rightarrow \infty} \left[\sum_{\lambda \in \Lambda_W \setminus \mathbf{P}} \frac{|\mathcal{F}_\mathbf{T}(\phi_k)(\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda \rangle|^{\tau q'}} \right]^{1/q'},$$

since we know that for $\lambda \in \mathbf{P}$ we get nothing. Finally, let us define $\varphi_k : \mathfrak{h}_\mathbb{R} \rightarrow \mathbb{C}$ by the relation $\phi_k = \varphi_k A_\delta$. The function φ_k satisfies $\varphi_k(W(x)) = \varphi_k(x)$ for all $W \in \mathcal{W}_G$ and is supported in $k^{-1}\mathfrak{T}$; hence we can understand φ_k as a central function on G . We can also say that, as a consequence of the well-known relation

$$(4) \quad A_\delta = \exp_{-\delta} \prod_{\alpha \in \mathcal{R}^+} (\exp_\alpha - 1),$$

φ_k has no singularities. Therefore Proposition 5.3 provides the following relation for some constant $\mathcal{B}(G, q)$ depending on G and q :

$$(5) \quad \|\widehat{I_\tau(f_0 A_\delta)}\|_{L^{q'}(\mathfrak{h}_\mathbb{R}^*)} = \mathcal{B}(G, q) \lim_{k \rightarrow \infty} \|\widehat{\varphi}_k\|_{\mathcal{L}^{q'}(\widehat{G})}.$$

On the other hand, since φ_k can be seen as a central function on G , we can estimate the norm of φ_k on $L^q(G)$. By the Weyl integration formula we get

$$\begin{aligned}\|\varphi_k\|_{L^q(G)} &= \left[\frac{1}{|\mathcal{W}_G|} \int_{\mathbf{T}} |\varphi_k A_\delta(t)|^q |A_\delta(t)|^{2-q} dm(t) \right]^{1/q} \\ &= \left[\frac{k^{\sigma q}}{|\mathcal{W}_G|} \int_{\mathfrak{X}} |f_0 A_\delta(kx)|^q |A_\delta(x)|^{2-q} dx \right]^{1/q} \\ &\leq \left[\frac{(2\pi)^{(2-q)|\mathcal{R}^+|}}{|\mathcal{W}_G|} k^{\sigma q} \int_{\mathfrak{X}} |f_0 A_\delta(kx)|^q \prod_{\alpha \in \mathcal{R}^+} |\alpha(x)|^{2-q} dx \right]^{1/q},\end{aligned}$$

where the last inequality follows from (4). Now, under the change of variable $y = kx$ and taking $\mathcal{C}(G, q) = (2\pi)^{\tau|\mathcal{R}^+|} |\mathcal{W}_G|^{-1/q}$, we obtain

$$\|\varphi_k\|_{L^q(G)} \leq \mathcal{C}(G, q) k^{\sigma - \tau|\mathcal{R}^+| - r/q} \left(\int_{\mathfrak{X}} |f_0 A_\delta(y)|^q \prod_{\alpha \in \mathcal{R}^+} |\alpha(y)|^{\tau q} dy \right)^{1/q}.$$

Recall that $\text{supp}(f_0 A_\delta) \subset \mathfrak{X}$; therefore the integral over $k\mathfrak{X}$ (the domain of integration after the change of variable) reduces to the same integral over \mathfrak{X} . But $\sigma - \tau|\mathcal{R}^+| - r/q = 0$, and the product inside the integral is bounded over \mathfrak{X} , say by M_G . Then we can write

$$(6) \quad \|\varphi_k\|_{L^q(G)} \leq \mathcal{C}(G, q) M_G \|f_0 A_\delta\|_{L^q(\mathfrak{h}_{\mathbb{R}})}.$$

In summary, by (5) and (6), we know there exists a constant $\mathcal{D}(G, q)$, depending on G and q , such that

$$\mathcal{K}(G, q) = \mathcal{D}(G, q) \frac{\|I_\tau(\widehat{f_0 A_\delta})\|_{L^{q'}(\mathfrak{h}_{\mathbb{R}}^*)}}{\|f_0 A_\delta\|_{L^q(\mathfrak{h}_{\mathbb{R}})}} \leq \liminf_{k \rightarrow \infty} \frac{\|\widehat{\varphi_k}\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|\varphi_k\|_{L^q(G)}} \leq 1.$$

Since f_0 is bounded, we easily obtain that $f_0 A_\delta \in L^q(\mathfrak{h}_{\mathbb{R}})$, $I_\tau(\widehat{f_0 A_\delta}) \in L^{q'}(\mathfrak{h}_{\mathbb{R}}^*)$ and $\mathcal{K}(G, q) > 0$. Therefore we have found a family $\{\varphi_k : k \geq 1\}$ of central functions on G whose supports are eventually in \mathcal{U} and such that their Hausdorff-Young quotient of exponent q is bounded below by a positive constant. This concludes the proof of Theorem 5.1 for compact semisimple simply connected Lie groups.

If G is not simply connected, some extra comments have to be made. We shall not give complete proofs of any of them; the details are left to the reader.

(i) Generalization (3) of formula (1) has no meaning here, but we can generalize formula (2) as

$$\widehat{f}(\pi_\lambda) = \frac{1}{d_\lambda} \det W \mathcal{F}_{\mathbf{T}}(f B_\delta)(W^t(\lambda + \delta) - \delta) 1_{d_\lambda}.$$

This generalization provides a couple of results parallel to Lemmas 5.2 and 5.5. Namely,

- We have $\{W^t(\lambda + \delta) - \delta : W \in \mathcal{W}_G, \lambda \in \Lambda_{\text{DW}}\} = \Lambda_{\mathbb{W}} \setminus (\text{P} - \delta)$. The mapping $(W, \lambda) \in \mathcal{W}_G \times \Lambda_{\text{DW}} \mapsto W^t(\lambda + \delta) - \delta \in \Lambda_{\mathbb{W}} \setminus (\text{P} - \delta)$ is injective.
- If $f : G \rightarrow \mathbb{C}$ is central, then $\mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f B_\delta)(\xi) = 0$ for all $\xi \in \text{P} - \delta$.

- (ii) Proposition 5.3 is now replaced by the following identity, valid for central functions $f : G \rightarrow \mathbb{C}$:

$$\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})} = \mathcal{A}(G, q) \left[\sum_{\lambda \in \Lambda_W \setminus (P-\delta)} \frac{|\mathcal{F}_T(fB_\delta)(\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda + \delta \rangle|^{q'-2}} \right]^{1/q'}.$$

- (iii) The bases of $\mathfrak{h}_{\mathbb{R}}^*$ and $\mathfrak{h}_{\mathbb{R}}$ respectively that generate Λ_W and L_W with integer coefficients are no longer the basis of fundamental weights and its predual. In fact, the fundamental weights generate the weight lattice of the universal covering group of G , which is a lattice containing Λ_W and strictly bigger than it. Therefore we need to define $\{H_1, H_2, \dots, H_r\}$ and $\{\omega_1, \omega_2, \dots, \omega_r\}$ just as the bases of $\mathfrak{h}_{\mathbb{R}}$ and $\mathfrak{h}_{\mathbb{R}}^*$ respectively for which L_W and Λ_W have integer coefficients. Once we have clarified this point, we can define \mathfrak{T} in the same way and regard f_0 as a bounded complex-valued function on $\mathfrak{h}_{\mathbb{R}}$, supported in \mathfrak{T} and symmetric under the reflections that generate \mathcal{W}_G .
- (iv) Let us recall that if $\delta \notin \Lambda_W$, the function A_δ is not well-defined on \mathbf{T} . But A_δ is originally defined on $\mathfrak{h}_{\mathbb{R}}$, and $\delta \notin \Lambda_W$ is not an obstacle to working with A_δ as a function defined on $\mathfrak{h}_{\mathbb{R}}$. On the other hand, (ii) leads us to consider, in the same spirit as in the proof given for simply connected groups, the function

$$\widetilde{I}_\tau(\widehat{f_0 B_\delta})(\xi) = \frac{1}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \xi + \delta \rangle|^\tau} \mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 B_\delta)(\xi).$$

Now, the remark about A_δ shows that $\widetilde{I}_\tau(\widehat{f_0 B_\delta})(\xi) = I_\tau(\widehat{f_0 A_\delta})(\xi + \delta)$. Hence we can proceed as before, expressing the norm of this function in $L^{q'}(\mathfrak{h}_{\mathbb{R}}^*)$ as a Riemann sum, but this time we take the lattice $\Lambda_W + \delta$ instead of Λ_W :

$$\|\widetilde{I}_\tau(\widehat{f_0 B_\delta})\|_{L^{q'}(\mathfrak{h}_{\mathbb{R}}^*)} = \lim_{k \rightarrow \infty} \left[\sum_{\lambda \in \Lambda_W + \delta} \frac{V_G}{k^r} \frac{|\mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 A_\delta)(k^{-1}\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, k^{-1}\lambda \rangle|^{\tau q'}} \right]^{1/q'}.$$

- (v) It is not difficult to check that $\mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 A_\delta)(k^{-1}\lambda) = k^{r-\sigma} \mathcal{F}_T(\varphi_k B_\delta)(\lambda - \delta)$, where φ_k is defined as above. Hence we get

$$\begin{aligned} \|\widetilde{I}_\tau(\widehat{f_0 B_\delta})\|_{L^{q'}(\mathfrak{h}_{\mathbb{R}}^*)} &= V_G^{1/q'} \lim_{k \rightarrow \infty} \left[\sum_{\lambda \in \Lambda_W \setminus (P-\delta)} \frac{|\mathcal{F}_T(\varphi_k B_\delta)(\lambda)|^{q'}}{\prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \lambda + \delta \rangle|^{\tau q'}} \right]^{1/q'} \\ &= \mathcal{B}(G, q) \lim_{k \rightarrow \infty} \|\widehat{\varphi}_k\|_{\mathcal{L}^{q'}(\widehat{G})}. \end{aligned}$$

Finally, to estimate the norm of φ_k on $L^q(G)$, we follow the same arguments. This completes the proof of Theorem 5.1 and, consequently, the proof of Theorem 3.1.

Remark 5.6. Let $\{\mathcal{U}_n : n \geq 1\}$ be a basis of neighborhoods of $\mathbf{1}$, and let

$$\mathcal{K}(G, q) = \inf_{n \geq 1} \sup \left\{ \frac{\|\widehat{f}\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|f\|_{L^q(G)}} : f \text{ central, } f \in L^q(G), \text{ supp}(f) \subset \mathcal{U}_n \right\}.$$

This constant does not depend on the chosen basis, and Theorem 5.1 states that $0 < \mathcal{K}(G, q) \leq 1$ for any $1 \leq q \leq 2$ and any compact semisimple Lie group. However, it would be interesting to find the exact value of that constant. Sharp constants for

the Hausdorff-Young inequality were investigated in [2], [3] and [13]. In the local case, if $\mathcal{B}_q = \sqrt{q^{1/q}/q'^{1/q'}}$ stands for the Babenko-Beckner constant, it is already known that $\mathcal{K}(\mathbb{T}, q) = \mathcal{B}_q$. Andersson proved it for q' an even integer in [1], and Sjölin completed the proof, see [15]. Also it is obvious that $\mathcal{K}(G, 1) = \mathcal{K}(G, 2) = 1$ for any compact group G . In the general case, a detailed look at the proof of Theorem 5.1 gives that the constant $\mathcal{K}(G, q)$ is the supremum of

$$|\mathcal{W}_G|^\tau \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \delta \rangle|^\tau V_G^{-1/q'} \lim_{k \rightarrow \infty} \frac{\left(\int_{\mathfrak{h}_{\mathbb{R}}^*} |\mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 A_\delta(\xi))|^{q'} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \xi \rangle|^{2-q'} d\xi \right)^{1/q'}}{\left(\int_{\mathfrak{h}_{\mathbb{R}}} |f_0 A_\delta(x)|^q |k|^{\mathcal{R}^+} |A_\delta(x/k)|^{2-q} dx \right)^{1/q}}$$

for $1 < q \leq 2$, where the supremum runs over the family of functions $f_0 : \mathfrak{h}_{\mathbb{R}} \rightarrow \mathbb{C}$, supported in \mathfrak{T} and symmetric under the reflections generating the Weyl group of G . If $\mathcal{K}_{f_0}(G, q)$ denotes the expression given above, then one easily gets that $\mathcal{K}_{f_0}(G, q)$ equals

$$\frac{|\mathcal{W}_G|^\tau}{(2\pi)^{\tau|\mathcal{R}^+|} V_G^{1/q'}} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \delta \rangle|^\tau \frac{\left(\int_{\mathfrak{h}_{\mathbb{R}}^*} |\mathcal{F}_{\mathfrak{h}_{\mathbb{R}}}(f_0 A_\delta(\xi))|^{q'} \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, \xi \rangle|^{2-q'} d\xi \right)^{1/q'}}{\left(\int_{\mathfrak{h}_{\mathbb{R}}} |f_0 A_\delta(x)|^q \prod_{\alpha \in \mathcal{R}^+} |\langle \alpha, x \rangle|^{2-q} dx \right)^{1/q}}.$$

Moreover, taking $q = 2$ and using the Plancherel theorem on compact groups, we see that $V_G = 1$. The boundedness of this expression can be regarded as a weighted Hausdorff-Young inequality of Pitt type; see [4] for more on this topic.

As we pointed out in the introduction, the growth of $\mathcal{C}_q^1(l^{p'}(n), G)$ remains open for $1 \leq p < q \leq 2$. We end this paper with some remarks about this problem.

Remark 5.7. In Theorem 3.1 we found an extremal function $\Phi_n = (\varphi_1, \varphi_2, \dots, \varphi_n)$ such that

$$\mathcal{C}_q^1(l^p(n), G) \geq \frac{\|\widehat{\Phi}_n\|_{\mathcal{L}_{l^{p'}(n)}^{q'}(\widehat{G})}}{\|\Phi_n\|_{L_{l^p(n)}^q(G)}} \geq \mathcal{K}(G, q) n^{1/p-1/q}.$$

Our functions $\varphi_1, \varphi_2, \dots, \varphi_n$ satisfied two crucial properties, namely,

- (P1) the norm of $\widehat{\varphi}_k(\pi)$ on $S_{d_\pi}^{q'}$ does not depend on k for any $\pi \in \widehat{G}$, and
- (P2) $\varphi_1, \varphi_2, \dots, \varphi_n$ have pairwise disjoint supports on G .

The idea was to compare the norms of $\widehat{\Phi}_n$ and Φ_n with $n^{1/p}$ and $n^{1/q}$ respectively. To this end, properties (P1) and (P2) were the conditions to be required, since they provided suitable simplifications for the original expressions of such norms. Now, if we replace $l^p(n)$ by $l^{p'}(n)$ in the relation above, we want to compare the norms of $\widehat{\Phi}_n$ and Φ_n with $n^{1/q'}$ and $n^{1/p'}$ respectively. Notice that $1/p - 1/q = 1/q' - 1/p'$. For that, we require these other properties on $\varphi_1, \varphi_2, \dots, \varphi_n$:

- (P3) the absolute value $|\varphi_k(g)|$ does not depend on k for any $g \in G$, and
- (P4) $\widehat{\varphi}_1, \widehat{\varphi}_2, \dots, \widehat{\varphi}_n$ have pairwise disjoint supports on \widehat{G} .

In the introduction we recalled that the growth of $\mathcal{C}_q^1(l^p(n), G)$ and $\mathcal{C}_q^1(l^{p'}(n), G)$ can be understood as dual problems with respect to the Fourier transform operator.

Now, these properties justify this point. Assuming properties (P3) and (P4), we get $\mathcal{C}_q^1(l^{p'}(n), G) \geq \mathcal{K}'(G, q, n)n^{1/q'-1/p'}$, where $\mathcal{K}'(G, q, n)$ is given by

$$\mathcal{K}'(G, q, n) = \left(\frac{1}{n} \sum_{k=1}^n \left[\frac{\|\widehat{\varphi}_k\|_{\mathcal{L}^{q'}(\widehat{G})}}{\|\varphi_k\|_{L^q(G)}} \right]^{q'} \right)^{1/q'}.$$

Hence, if we define $\mathcal{K}'(G, q) = \inf_{n \geq 1} \mathcal{K}'(G, q, n)$, it remains to see that $\mathcal{K}'(G, q) > 0$. We do not know if this inequality holds for any compact semisimple Lie group and any $1 \leq q \leq 2$.

Remark 5.8. We do not know if properties (P3) and (P4) are compatible. However, given $f_0 \in L^2(G)$ continuous and any sequence of positive numbers $\{\varepsilon_n : n \geq 1\}$ decreasing to 0, it is not difficult to see that there exists a system $\Phi = \{\varphi_n : n \geq 1\}$ of trigonometric polynomials on G satisfying the following three conditions:

- (1) The functions $\widehat{\varphi}_1, \widehat{\varphi}_2, \dots$ have pairwise disjoint supports on \widehat{G} .
- (2) The estimate $|\varphi_n| \leq |f_0| + \varepsilon_n$ holds in G .
- (3) The estimate $|\varphi_n| \geq |f_0| - \varepsilon_n$ holds outside Ω_n , where $\mu(\Omega_n) \rightarrow 0$ as $n \rightarrow \infty$.

Remark 5.9. As is well known, $\mathcal{C}_q^1(l^{p'}(n), G) = n^{1/p-1/q}$ for any compact abelian group G . This equality follows by taking $\varphi_1, \varphi_2, \dots, \varphi_n$ to be a collection of n pairwise distinct characters. This motivates us to see what happens when we consider the irreducible characters of a compact semisimple Lie group. Let χ_λ be the character of the irreducible representation π_λ , and consider the function $\Phi_n(g) = (d_{\lambda_1}^\tau \chi_{\lambda_1}(g), d_{\lambda_2}^\tau \chi_{\lambda_2}(g), \dots, d_{\lambda_n}^\tau \chi_{\lambda_n}(g))$, where $\lambda_1, \lambda_2, \dots, \lambda_n$ are pairwise distinct dominant weights and $\tau = 1 - 2/q'$. Then we have

$$\|\widehat{\Phi}_n\|_{\mathcal{L}_{l^{p'}(n)}^{q'}(\widehat{G})} = \left(\sum_{k=1}^n d_{\lambda_k} \|d_{\lambda_k}^\tau \widehat{\chi}_{\lambda_k}(\pi_k)\|_{S_{d_{\lambda_k}}^{q'}}^{q'} \right)^{1/q'} = n^{1/q'}.$$

On the other hand, applying consecutively the Weyl integration formula and the Weyl character formula, we get

$$\|\Phi_n\|_{L_{l^{p'}(n)}^q(G)} = \left(\frac{1}{|\mathcal{W}_G|} \int_{\mathbf{T}} \left(\sum_{k=1}^n |d_{\lambda_k} A_\delta(t)|^{\tau p'} |A_{\lambda_k + \delta}(t)|^{p'} \right)^{q/p'} dm(t) \right)^{1/q}.$$

However, these relations do not provide optimal growth. For instance, in the simplest case $G = SU(2)$ it can be checked that there exists a constant $\mathcal{K}_{p,q}$, depending on p and q , such that

$$\frac{1}{n^{1/p'}} \left(\int_{SU(2)} \|\Phi_n(g)\|_{l^{p'}(n)}^q d\mu(g) \right)^{1/q} \geq \mathcal{K}_{p,q} n^\tau.$$

Remark 5.10. If we try to find out why our attempts to get optimal growth have failed, we need to revisit the proof of Theorem 3.1. The point is that we required the functions $\varphi_1, \varphi_2, \dots, \varphi_n$ not only to satisfy properties (P1) and (P2), but also to be translations of a common function. This was essential in Section 3, and here the obstacle lies in the fact that we cannot take translations, since the dual object does not have a group structure. This is the main difference from the abelian case

where, since the dual object is a group, multiplication by a character in G becomes a translation in the other side of the Fourier transform operator.

Remark 5.11. The quantized Rademacher system associated to a probability space $(\Omega, \mathcal{M}, \mu)$, an index set Σ , and a family $\{d_\sigma : \sigma \in \Sigma\}$ of positive integers is defined by a collection $\mathcal{R} = \{\rho^\sigma : \Omega \rightarrow O(d_\sigma)\}_{\sigma \in \Sigma}$ of independent random orthogonal matrices, uniformly distributed on the orthogonal group $O(d_\sigma)$. In [9] we define the notions of \mathcal{R} -type, \mathcal{R} -cotype and strong \mathcal{R} -cotype of an operator space E . Moreover, we show that

$$\begin{aligned} \text{Fourier type } p &\Rightarrow \text{strong } \mathcal{R}\text{-cotype } p', \\ \text{Fourier cotype } p' &\Rightarrow \mathcal{R}\text{-type } p. \end{aligned}$$

These implications allow us to work with the quantized Rademacher system, where other techniques are available to study the growth of $\mathcal{C}_q^1(l^{p'}(n), G)$.

Remark 5.12. Of course, the growth of $\mathcal{C}_q^1(l^{p'}(n), G)$ is trivially optimal when we work with compact groups with infinitely many inequivalent irreducible representations of the same degree d_0 . The unitary groups $U(n)$ are the simplest non-commutative examples of this degenerate case. Also, it is not difficult to check that $\mathcal{C}_2^1(l^{p'}(n), G) = n^{1/2-1/p'}$ by the Plancherel theorem for compact groups.

REFERENCES

1. M. E. Andersson, *The Hausdorff-Young inequality and Fourier type*, Ph.D. Thesis, Uppsala (1993).
2. K. I. Babenko, *An inequality in the theory of Fourier integrals*, *Izv. Akad. Nauk SSSR* **25** (1961), 531 – 542; English transl., *Amer. Math. Soc. Transl. (2)* **44** (1965), 115 – 128. MR **25**:2379
3. W. Beckner, *Inequalities in Fourier analysis*, *Ann. of Math. (2)* **102** (1975), 159 – 182. MR **52**:6317
4. W. Beckner, *Pitt's inequality and the uncertainty principle*, *Proc. Amer. Math. Soc.* **123** (1995), 1897 – 1905. MR **95g**:42021
5. E. G. Effros and Z. J. Ruan, *Operator Spaces*, *London Math. Soc. Monogr.* **23**, Oxford Univ. Press (2000). MR **2002a**:46082
6. G. B. Folland, *A Course in Abstract Harmonic Analysis*, *Stud. Adv. Math.*, CRC Press (1995). MR **98c**:43001
7. W. Fulton and J. Harris, *Representation Theory: A First Course*, *Graduate Texts in Math.*, Springer-Verlag, New York, 1991. MR **93a**:20069
8. J. García-Cuerva and J. Parcet, *Vector-valued Hausdorff-Young inequality on compact groups*, Submitted for publication.
9. J. García-Cuerva and J. Parcet, *Quantized orthonormal systems: A non-commutative Kwapien theorem*, *Studia Math.* **155** (2003), 273 – 294.
10. R. A. Kunze, *L_p Fourier transforms on locally compact unimodular groups*, *Trans. Amer. Math. Soc.* **89** (1958), 519 – 540. MR **20**:6668
11. G. Pisier, *The Operator Hilbert Space OH , Complex Interpolation and Tensor Norms*, *Mem. Amer. Math. Soc.* **122** (1996), No. 585. MR **97a**:46024
12. G. Pisier, *Non-commutative vector valued L_p -spaces and completely p -summing maps*, *Astérisque (Soc. Math. France)* **247** (1998). MR **2000a**:46108
13. B. Russo, *The norm of the L^p -Fourier transform on unimodular groups*, *Trans. Amer. Math. Soc.* **192** (1974), 293 – 305. MR **55**:8689a

14. B. Simon, *Representations of Finite and Compact Groups*, Graduate Stud. Math. **10**, Amer. Math. Soc., Providence, RI, 1996. MR **97c**:22001
15. P. Sjölin, *A remark on the Hausdorff-Young inequality*, Proc. Amer. Math. Soc. **123** (1995), 3085 – 3088. MR **95m**:42007

DEPARTMENT OF MATHEMATICS, UNIVERSIDAD AUTÓNOMA DE MADRID, MADRID 28049, SPAIN
E-mail address: jose.garcia-cuerva@uam.es

DEPARTMENT OF MATHEMATICS, UNIVERSIDAD AUTÓNOMA DE MADRID, MADRID 28049, SPAIN

DEPARTMENT OF MATHEMATICS, UNIVERSIDAD AUTÓNOMA DE MADRID, MADRID 28049, SPAIN
E-mail address: javier.parcet@uam.es

LEFT-DETERMINED MODEL CATEGORIES
AND UNIVERSAL HOMOTOPY THEORIES

J. ROSICKÝ AND W. THOLEN

ABSTRACT. We say that a model category is left-determined if the weak equivalences are generated (in a sense specified below) by the cofibrations. While the model category of simplicial sets is not left-determined, we show that its non-oriented variant, the category of symmetric simplicial sets (in the sense of Lawvere and Grandis) carries a natural left-determined model category structure. This is used to give another and, as we believe simpler, proof of a recent result of D. Dugger about universal homotopy theories.

1. INTRODUCTION

Recall that a *model category* \mathcal{K} is a complete and cocomplete category \mathcal{K} equipped with three classes of morphisms \mathcal{C} , \mathcal{W} and \mathcal{F} , called *cofibrations*, *weak equivalences* and *fibrations*, such that

- (1) $(\mathcal{C}, \mathcal{F} \cap \mathcal{W})$ and $(\mathcal{C} \cap \mathcal{W}, \mathcal{F})$ are weak factorization systems and
- (2) \mathcal{W} is closed under retracts (in the category $\mathcal{K}^{\rightarrow}$ of morphisms of \mathcal{K}) and has the 2-out-of-3 property

(see [Q], [H], [Ho] or [AHRT2]). Model categories were introduced by D. Quillen to provide a foundation of homotopy theory. Here a weak factorization system is a pair $(\mathcal{L}, \mathcal{R})$ of morphisms such that every morphism has a factorization as an \mathcal{L} -morphism followed by an \mathcal{R} -morphism, and $\mathcal{R} = \mathcal{L}^{\square}$, $\mathcal{L} = {}^{\square}\mathcal{R}$ where \mathcal{L}^{\square} (${}^{\square}\mathcal{R}$) consists of all morphisms having the right (left) lifting property w.r.t. \mathcal{L} (\mathcal{R} , respectively). The morphism l has a *left lifting property* with respect to a morphism r (or r has a *right lifting property* w.r.t. l) if in every commutative square

$$\begin{array}{ccc} A & \xrightarrow{u} & C \\ \downarrow l & & \downarrow r \\ B & \xrightarrow{v} & D \end{array}$$

there exists a diagonal $d : B \rightarrow C$.

Received by the editors June 1, 2002.
2000 *Mathematics Subject Classification*. Primary 55U35.
The first author was supported by the Grant Agency of the Czech Republic under Grant 201/99/0310. The hospitality of the York University is gratefully acknowledged.
The second author was supported by the Natural Sciences and Engineering Council of Canada.

A model category is determined by any two of the three classes above. Clearly, \mathcal{C} and \mathcal{W} determine \mathcal{F} because $\mathcal{F} = (\mathcal{C} \cap \mathcal{W})^\square$, and from \mathcal{C} and \mathcal{F} one obtains the morphisms of \mathcal{W} as composites $g \cdot f$ with $f \in {}^\square\mathcal{F}$ and $g \in \mathcal{C}^\square$. In this paper we are interested in the model categories whose model structure is determined by its cofibrations only, and we therefore call them *left-determined*. For example, the model category **SComp** of simplicial complexes is left-determined while the model category **Simp** of simplicial sets is not left-determined.

Simp is, of course, the presheaf category $\mathbf{Set}^{\Delta^{op}}$ where Δ is the category of nonzero finite ordinals and order-preserving maps. F. W. Lawvere [L] and M. Grandis [G] introduced *symmetric simplicial sets* as functors $\mathbf{F}^{op} \rightarrow \mathbf{Set}$ where \mathbf{F} is the category of nonzero finite cardinals (= finite sets) and arbitrary maps. We will show that the category **SSimp** = $\mathbf{Set}^{\mathbf{F}^{op}}$ of symmetric simplicial sets is a left-determined model category. Moreover, the model categories **SSimp** and **Simp** are Quillen equivalent, i.e., they have equivalent homotopy categories.

D. Dugger [D] has recently shown that, for a small category \mathcal{X} , $\mathbf{Simp}^{\mathcal{X}^{op}}$ is a universal model category over \mathcal{X} . In particular, **Simp** is a universal model category over the (one-morphism) category **1**. We will give another proof of his result, by showing that also **SSimp** $^{\mathcal{X}^{op}}$ serves as a universal model category over \mathcal{X} . Since **SSimp** is left-determined, our proof is simpler.

The first author would like to acknowledge his gratitude to Tibor Beke for introducing him to homotopy theory. Both authors acknowledge stimulating conversations with him regarding the subject of this paper.

After having completed this work we learned that the concept of a left-determined model category was independently developed by J. H. Smith [S], who used the term *minimal model category* instead. He also observed that the usual model structure on simplicial sets fails to be left-determined.

2. LEFT-DETERMINED MODEL CATEGORIES

Definition 2.1. A model category \mathcal{K} is *left-determined* if \mathcal{W} is the smallest class of morphisms satisfying the following conditions:

- (i) $\mathcal{C}^\square \subseteq \mathcal{W}$,
- (ii) \mathcal{W} is closed under retracts and satisfies the 2-out-of-3 property,
- (iii) $\mathcal{C} \cap \mathcal{W}$ is stable under pushout and closed under transfinite composition.

We will denote the smallest class of morphisms satisfying (i)–(iii) by $\mathcal{W}_{\mathcal{C}}$. It has the property that $\mathcal{W}_{\mathcal{C}} \subseteq \mathcal{W}$ for each model category \mathcal{K} having \mathcal{C} as the class of cofibrations. Left-determined model categories are those for which $\mathcal{W} = \mathcal{W}_{\mathcal{C}}$. Recall that \mathcal{C}^\square denotes the class of morphisms having the right lifting property w.r.t. \mathcal{C} . Of course, $\mathcal{C}^\square = \mathcal{F} \cap \mathcal{W}$ is the class of trivial fibrations.

In general, given \mathcal{C} , the first principal problem is whether \mathcal{C} and $\mathcal{W}_{\mathcal{C}}$ yield a model category. The next theorem gives an affirmative answer under an additional set-theoretic hypothesis, the Vopěnka's Principle. (Subsequently, J. H. Smith informed us that he has been able to prove the theorem even absolutely, i.e., without any additional set-theoretic hypothesis.) Recall that Vopěnka's Principle is a set-theoretic axiom implying the existence of very large cardinals (see [AR]). We denote by $\text{cof}(\mathcal{I})$ the smallest class of morphisms containing \mathcal{I} , closed under retracts in comma-categories $A \backslash \mathcal{K}$ and satisfying (iii). The smallest class containing \mathcal{I} and satisfying (iii) is denoted by $\text{cell}(\mathcal{I})$ (see [AHRT1]).

Theorem 2.2. *Let \mathcal{I} be a (small) set of morphisms in a locally presentable category \mathcal{K} . Under Vopěnka's principle, $\mathcal{C} = \text{cof}(\mathcal{I})$ and $\mathcal{W} = \mathcal{W}_{\mathcal{C}}$ yield a model category structure on \mathcal{K} .*

Proof. According to the theorem of J. H. Smith (see [B, 1.7]), it suffices to show that $\mathcal{W}_{\mathcal{C}}$ satisfies the solution set condition at \mathcal{I} . This means that for every $f \in \mathcal{I}$ there is a subset \mathcal{X}_f of $\mathcal{W}_{\mathcal{C}}$ such that every morphism $f \rightarrow g$, $g \in \mathcal{W}_{\mathcal{C}}$, factorizes through some $h \in \mathcal{X}_f$. Since $f \backslash \mathcal{W}_{\mathcal{C}}$ is a full subcategory of $f \backslash \mathcal{K}$ and $f \backslash \mathcal{K}$ is locally presentable (see [AR, 1.57]), $f \backslash \mathcal{W}_{\mathcal{C}}$ has a small dense subcategory \mathcal{X}_f provided that we assume Vopěnka's principle (see [AR, 6.6]). Without any loss of generality, we may assume that \mathcal{X}_f contains the initial object of $f \backslash \mathcal{W}_{\mathcal{C}}$ provided that it exists. A morphism $f \rightarrow g$ in $f \backslash \mathcal{W}_{\mathcal{C}}$ is either initial in $f \backslash \mathcal{W}_{\mathcal{C}}$ and thus belongs to \mathcal{X}_f , or it factorizes through some morphism $f \rightarrow h$ from \mathcal{X}_f . Hence \mathcal{X}_f , $f \in \mathcal{I}$, provide a solution set condition at \mathcal{I} . \square

A model category is called *cofibrantly generated* if $\mathcal{C} = \text{cof}(\mathcal{I})$ and $\mathcal{C} \cap \mathcal{W} = \text{cof}(\mathcal{J})$ for sets \mathcal{I} and \mathcal{J} . Following J. H. Smith, a model category \mathcal{K} is called *combinatorial* if it is cofibrantly generated and the category \mathcal{K} is locally presentable. The model categories from Theorem 2.2 are combinatorial.

Left-determined model categories are, in some sense, related to left Bousfield localizations. Recall that, having model categories \mathcal{K} and \mathcal{L} , a *left Quillen functor* $H : \mathcal{K} \rightarrow \mathcal{L}$ is a left adjoint functor preserving cofibrations and trivial cofibrations (i.e., elements of $\mathcal{C} \cap \mathcal{W}$). Every left Quillen functor preserves weak equivalences between cofibrant objects (see [Ho]). An object A of a model category \mathcal{K} is *cofibrant* if $0 \rightarrow A$ is a cofibration.

A model category \mathcal{K} is called *functorial* if both weak factorization systems $(\mathcal{C}, \mathcal{F} \cap \mathcal{W})$ and $(\mathcal{C} \cap \mathcal{W}, \mathcal{F})$ are functorial. This means that, for a weak factorization system $(\mathcal{L}, \mathcal{R})$, there is a functor $F : \mathcal{K}^{\rightarrow} \rightarrow \mathcal{K}$ and natural transformations $\lambda : \text{dom} \rightarrow F$ and $\varrho : F \rightarrow \text{cod}$ such that $f = \varrho_f \cdot \lambda_f$ is an $(\mathcal{L}, \mathcal{R})$ -factorization of a morphism $f : A \rightarrow B$; of course, $\text{dom}(f) = A$ and $\text{cod}(f) = B$. This definition of a functorial weak factorization system is given in [RT] where its relation to functoriality in the sense of Hovey [Ho] is explained. Each combinatorial model category is functorial. In a functorial model category \mathcal{K} we have a *cofibrant replacement functor* $Q : \mathcal{K} \rightarrow \mathcal{K}$ where

$$0 \longrightarrow Q(A) \xrightarrow{q_A} A$$

is a functorial weak factorization in $(\mathcal{C}, \mathcal{F} \cap \mathcal{W})$. Then $q : Q \rightarrow \text{Id}_{\mathcal{K}}$ is a natural transformation.

Let \mathcal{K} be a model category and \mathcal{Z} a class of morphisms of \mathcal{K} . A *left Bousfield localization* of \mathcal{K} w.r.t. \mathcal{Z} is a model category structure $\mathcal{K} \backslash \mathcal{Z}$ on the category \mathcal{K} such that

- (a) $\mathcal{K} \backslash \mathcal{Z}$ has the same cofibrations as \mathcal{K} ,
- (b) weak equivalences of $\mathcal{K} \backslash \mathcal{Z}$ contain both the weak equivalences of \mathcal{K} and the morphisms of \mathcal{Z} and
- (c) each left Quillen functor $H : \mathcal{K} \rightarrow \mathcal{L}$ such that $H \cdot Q$ sends \mathcal{Z} -morphisms to weak equivalences is a left Quillen functor $\mathcal{K} \backslash \mathcal{Z} \rightarrow \mathcal{L}$

(see [H, 3.3.1]). J. H. Smith proved that if \mathcal{K} is a left proper combinatorial model category and \mathcal{Z} is a set of morphisms, then a left Bousfield localization $\mathcal{K} \backslash \mathcal{Z}$ exists (see [S]). (The model category is called *left proper* if every pushout of a weak

equivalence along a cofibration is a weak equivalence.) As a consequence, we get the following result.

Theorem 2.3. *Let \mathcal{K} be a left proper, combinatorial model category and \mathcal{Z} a class of morphisms of \mathcal{K} . Under Vopěnka’s principle, a left Bousfield localization $\mathcal{K}\backslash\mathcal{Z}$ exists.*

Proof. We can express \mathcal{Z} as a union of an increasing chain of (small) subsets \mathcal{Z}_i indexed by ordinals. Let \mathcal{W}_i denote the class of weak equivalences in the model category $\mathcal{K}\backslash\mathcal{Z}_i$ (which exists by the result of J. Smith). Then we have $\mathcal{W}_i \subseteq \mathcal{W}_j$ for $i \leq j$; this follows from $Id_{\mathcal{K}} : \mathcal{K} \rightarrow \mathcal{K}\backslash\mathcal{Z}_j$ being a left Quillen functor sending \mathcal{Z}_i morphisms to weak equivalences. Hence $Q = Id_{\mathcal{K}} \cdot Q : \mathcal{K}\backslash\mathcal{Z}_i \rightarrow \mathcal{K}\backslash\mathcal{Z}_j$ is a left Quillen functor. Let $f : A \rightarrow B$ be a weak equivalence from \mathcal{W}_i and consider

$$\begin{array}{ccc} QA & \xrightarrow{Qf} & QB \\ \downarrow r_A & & \downarrow r_B \\ A & \xrightarrow{f} & B \end{array}$$

Since $r_A, r_B \in \mathcal{C}^\square$, we have $r_A, r_B, Qf \in \mathcal{W}_j$. Hence $f \in \mathcal{W}_j$. Put $\mathcal{W}_* = \bigcup_{i \in \text{Ord}} \mathcal{W}_i$. Then \mathcal{W}_* is closed under retracts and satisfies the 2-out-of-3 property. Analogously as in Theorem 2.2, Vopěnka’s principle guarantees that \mathcal{K}, \mathcal{C} and \mathcal{W}_* form a model category $\mathcal{K}\backslash\mathcal{Z}$.

Let $H : \mathcal{K} \rightarrow \mathcal{L}$ be a left Quillen functor such that $H \cdot Q$ sends \mathcal{Z} -morphisms to weak equivalences. Since $H : \mathcal{K}\backslash\mathcal{Z} \rightarrow \mathcal{L}$ is a left Quillen functor for each i , $H : \mathcal{K}\backslash\mathcal{Z}_i \rightarrow \mathcal{L}$ is a left Quillen functor. Hence $\mathcal{K}\backslash\mathcal{Z}$ is a left Bousfield localization of \mathcal{K} w.r.t. \mathcal{Z} . □

Using [AR] as well, C. Casacuberta, D. Sceveneles and J. H. Smith [CSS] proved a related result saying that cohomological localizations of simplicial sets exist under Vopěnka’s Principle.

Remark 2.4. In analogy with the definition of a left-determined model category, we define $\mathcal{W}_{\mathcal{X}}$, where \mathcal{X} is a class of morphisms in a model category \mathcal{K} , as the smallest class of morphisms satisfying

- (i) $\mathcal{W} \cup \mathcal{X} \subseteq \mathcal{W}_{\mathcal{X}}$,
- (ii) $\mathcal{W}_{\mathcal{X}}$ is closed under retracts and satisfies the 2-out-of-3 property,
- (iii) $\mathcal{C} \cap \mathcal{W}_{\mathcal{X}}$ is stable under pushout and closed under transfinite composition.

Under Vopěnka’s principle, \mathcal{K}, \mathcal{C} and $\mathcal{W}_{\mathcal{X}}$ form a model category structure for each combinatorial model category \mathcal{K} . This model category structure is evidently $\mathcal{K}\backslash\mathcal{X}$, provided that $\mathcal{K}\backslash\mathcal{X}$ exists (because $\mathcal{W}_{\mathcal{X}}$ is contained in the class of weak equivalences of $\mathcal{K}\backslash\mathcal{X}$).

Let \mathcal{K} be a cofibrantly generated model category and \mathcal{X} a small category. Then there is a cofibrantly generated model category structure on the functor category $\mathcal{K}^{\mathcal{X}^{op}}$ (see [H, 14.2.1]); this structure is called the *Bousfield-Kan structure*. To recall it we denote by

$$ev_{\mathcal{X}} : \mathcal{K}^{\mathcal{X}^{op}} \rightarrow \mathcal{K}$$

the evaluation functor given by $ev_X(A) = A(X)$ and by

$$F_X : \mathcal{K} \rightarrow \mathcal{K}^{\mathcal{X}^{op}}$$

its left adjoint given by

$$F_X(K)(Y) = \coprod_{\mathcal{X}^{op}(X,Y)} K.$$

If \mathcal{I} (resp., \mathcal{J}) is the set of generating (resp., trivial) cofibrations in \mathcal{K} , then the Bousfield-Kan model structure has $\overline{\mathcal{I}} = \bigcup_{X \in \text{ob}(\mathcal{X})} F_X(\mathcal{I})$ as generating cofibrations

and $\overline{\mathcal{J}} = \bigcup_{X \in \text{ob}(\mathcal{X})} F_X(\mathcal{J})$ as generating trivial cofibrations. Then

- (a) $\varphi : A \rightarrow B$ is a weak equivalence in $\mathcal{K}^{\mathcal{X}^{op}}$ iff $\varphi_X : A(X) \rightarrow B(X)$ is a weak equivalence in \mathcal{K} for each X in \mathcal{X} ,
- (b) $\varphi : A \rightarrow B$ is a fibration in $\mathcal{K}^{\mathcal{X}^{op}}$ iff $\varphi_X : A(X) \rightarrow B(X)$ is a fibration in \mathcal{K} for each X in \mathcal{X} .

Consequently, trivial fibrations are also morphisms in $\mathcal{K}^{\mathcal{X}^{op}}$, which are pointwise trivial fibrations in \mathcal{K} .

Proposition 2.5. *Let \mathcal{K} be a cofibrantly generated, left-determined model category and \mathcal{X} a small category. Then $\mathcal{K}^{\mathcal{X}^{op}}$ is a left-determined model category.*

Proof. Let \mathcal{I} (resp., \mathcal{J}) be the set of generating (resp., trivial) cofibrations in \mathcal{K} , and let $\overline{\mathcal{W}}$ be the set of weak equivalences in $\mathcal{K}^{\mathcal{X}^{op}}$. We have $\mathcal{W}_{\text{cof}(\overline{\mathcal{I}})} \subseteq \overline{\mathcal{W}}$. Let $w \in \overline{\mathcal{W}}$. Then $w = f \cdot g$ where $f \in \overline{\mathcal{I}}^{\square}$ and $g \in \text{cof}(\overline{\mathcal{J}})$. We have $f \in \mathcal{W}_{\text{cof}(\overline{\mathcal{I}})}$. To prove that $g \in \mathcal{W}_{\text{cof}(\overline{\mathcal{I}})}$, that is, $w \in \mathcal{W}_{\text{cof}(\overline{\mathcal{I}})}$, it suffices to show that $\overline{\mathcal{J}} \subseteq \mathcal{W}_{\text{cof}(\overline{\mathcal{I}})}$. But this follows from $\mathcal{J} \subseteq \mathcal{W}_{\text{cof}(\mathcal{I})}$ and the fact that F_X preserve colimits. \square

3. SYMMETRIC SIMPLICIAL SETS

A trivial example of a left-determined model category is the category **Set** of sets, with \mathcal{C} the class of all monomorphisms, and with \mathcal{W} the class of the morphisms between nonempty sets and the identity morphism on \emptyset . To give a nontrivial example we recall that a *simplicial complex* is a set X equipped with a set \mathcal{X} of nonempty finite subsets of X such that

- (a) $\{x\} \in \mathcal{X}$ for each $x \in X$,
- (b) $A \in \mathcal{X}, \emptyset \neq B \subseteq A \Rightarrow B \in \mathcal{X}$.

Elements $A \in \mathcal{X}$ with $|A| = n + 1$ are called (non-degenerated) *n-simplices*. For $n = 0, 1$ and 2 we speak about *vertices*, *edges* and *triangles*, respectively. If $|A| \leq 2$ for each $A \in \mathcal{X}$, then (X, \mathcal{X}) is called a (non-oriented) *graph* (with loops). Morphisms of complexes $(X, \mathcal{X}) \rightarrow (Y, \mathcal{Y})$ are maps $h : X \rightarrow Y$ with $h(\mathcal{X}) \subseteq \mathcal{Y}$. We denote the category of simplicial complexes by **SComp**. We will show that $\mathcal{C} = \text{Mono}$ yields a left-determined model category structure on **SComp**. But the disadvantage of simplicial complexes is that each simplex is uniquely determined by its vertices, which makes colimits in **SComp** bad. This led S. Eilenberg and B. Zilber [EZ] to introduce complete semisimplicial complexes, which later were renamed as simplicial sets, and which are oriented. Surprisingly, non-oriented simplicial sets were introduced only recently by F. W. Lawvere [L] and M. Grandis [G]; they are called symmetric simplicial sets. Their position to simplicial complexes

is the same as the position of multigraphs to graphs in graph theory (one admits multiple edges).

Definition 3.1. Let \mathbf{F} denote the category of nonzero finite cardinals (and all maps). A *symmetric simplicial set* is by definition a functor $\mathbf{F}^{\text{op}} \rightarrow \mathbf{Set}$. The category $\mathbf{Set}^{\mathbf{F}^{\text{op}}}$ of symmetric simplicial sets will be denoted by \mathbf{SSimp} .

We will recall the basic properties of symmetric simplicial sets (see [G]). We have the Yoneda embedding

$$Y : \mathbf{F} \rightarrow \mathbf{SSimp}.$$

Its values $\Delta_{n-1} = Y(n)$ are in fact simplicial complexes, which yields the functor

$$\mathbf{F} \rightarrow \mathbf{SComp}.$$

The Yoneda embedding Y extends along this functor to the full embedding

$$G : \mathbf{SComp} \rightarrow \mathbf{SSimp}$$

sending a simplicial complex (X, \mathcal{X}) to the functor $A : \mathbf{F}^{\text{op}} \rightarrow \mathbf{Set}$ given by $A(n) = \{S \in \mathcal{X} \mid |S| \leq n-1\}$. In what follows we will identify a simplicial complex (X, \mathcal{X}) with its image under G . Hence \mathbf{SComp} will be considered as a full subcategory of \mathbf{SSimp} .

We have the functor $U : \mathbf{SSimp} \rightarrow \mathbf{Set}$ given by precomposition with $\mathbf{1} \rightarrow \mathbf{F}^{\text{op}}$ (sending the object of $\mathbf{1}$ to 1). We can view $U(A)$ as the set of vertices of a symmetric simplicial set A and the whole A as the set $U(A)$ equipped with n -simplices corresponding to morphisms $\Delta_n \rightarrow A$. We will use the notation $A = (UA, \mathcal{A})$ where \mathcal{A} is the set of simplices of A . For instance, Δ_n has all nonempty subsets of $\{0, 1, \dots, n\}$ as simplices. But note that the functor U is not faithful.

The embedding $\Delta \rightarrow \mathbf{F}$ induces the faithful functor

$$H : \mathbf{SSimp} \rightarrow \mathbf{Simp}.$$

It has a left adjoint

$$L : \mathbf{Simp} \rightarrow \mathbf{SSimp}$$

sending each simplicial set to its symmetrization. There is also a right adjoint

$$R : \mathbf{Simp} \rightarrow \mathbf{SSimp}.$$

Let $\partial\Delta_n$ be the boundary of Δ_n for $n > 0$, i.e., $U(\partial\Delta_n) = n+1$, and simplices of $\partial\Delta_n$ are all nonempty subsets of $\{0, 1, \dots, n\}$ distinct from $\{0, 1, \dots, n\}$. Let $i_n : \partial\Delta_n \rightarrow \Delta_n$, $n > 0$, be the embeddings. Let $\mathcal{I} = \{i_n \mid n \geq 0\}$ where

$$i_0 : 0 \rightarrow \Delta_0.$$

In what follows, the class of all monomorphisms of \mathbf{SSimp} is denoted by $Mono$.

Lemma 3.2. $\text{cof}(\mathcal{I}) = Mono$.

The proof is the same as for simplicial sets.

Given $0 \leq k \leq n$, the k -horn Δ_n^k is the simplicial complex whose simplices are all subsets $\emptyset \neq S \subsetneq \{0, 1, \dots, n\}$ distinct from $\{0, \dots, k-1, k+1, \dots, n\}$. Let \mathcal{J} be the set of inclusions

$$j_n : \Delta_n^0 \rightarrow \Delta_n, \quad n > 0.$$

Lemma 3.3. $j_n \in \mathcal{W}_{Mono} \cap Mono$ for each $n > 0$.

Proof. Evidently, each morphism $s_n : \Delta_n \rightarrow \Delta_0$, $n \geq 0$, belongs to $Mono^\square$. Therefore, by 2.1 (i) and (ii), each morphism $u : \Delta_0 \rightarrow \Delta_n$ belongs to \mathcal{W}_{Mono} and, consequently, to $\mathcal{W}_{Mono} \cap Mono$. Hence $j_1 \in \mathcal{W}_{Mono} \cap Mono$.

Assume that $j_1, \dots, j_n \in \mathcal{W}_{Mono} \cap Mono$. Consider the pushout

$$\begin{array}{ccc} \Delta_n + \Delta_0 & \xrightarrow{p_n} & \Delta_n \\ \text{id}_{\Delta_n} + u_1^0 \downarrow & & \downarrow g_n \\ \Delta_n + \Delta_1 & \longrightarrow & P_n \end{array}$$

where $u_1^0(0) = 0$ and p_n is induced by id_{Δ_n} and $u_n^0 : \Delta_0 \rightarrow \Delta_n$ (again $u_n^0(0) = 0$). Since $\text{id}_{\Delta_n} + u_1^0 \in \mathcal{W}_{Mono} \cap Mono$, we have $g_n \in \mathcal{W}_{Mono} \cap Mono$ (by 2.1 (iii)). P_n is the simplicial complex given by attaching an edge at the vertex 0. By successive use of j_2, \dots, j_n , we fill horns to simplices. This is done via pushouts, starting with

$$\begin{array}{ccc} \Delta_2^0 & \xrightarrow{h} & P_n \\ j_2 \downarrow & & \downarrow \\ \Delta_2 & \longrightarrow & P'_n \end{array}$$

where h sends one edge of Δ_2^0 to the attached edge and the other edge to an edge of Δ_n . Doing this for all edges of Δ_n containing 0, we start to fill by using j_3 , etc. At the end we obtain Δ_{n+1}^0 and a morphism

$$q_n : \Delta_n \xrightarrow{g_n} P_n \longrightarrow P'_n \longrightarrow \dots \longrightarrow \Delta_{n+1}^0$$

which belongs to $\mathcal{W}_{Mono} \cap Mono$. Since, in the diagram

$$\begin{array}{ccc} \Delta_n & \xrightarrow{q_n} & \Delta_{n+1}^0 \\ s_n \searrow & & \swarrow t_n \\ & \Delta_0 & \end{array}$$

we have $s_n \in \mathcal{W}_{Mono}$, we get $t_n \in \mathcal{W}_{Mono}$. Since, in the diagram

$$\begin{array}{ccc} \Delta_{n+1}^0 & \xrightarrow{j_{n+1}} & \Delta_{n+1} \\ t_n \searrow & & \swarrow s_{n+1} \\ & \Delta_0 & \end{array}$$

we have $t_n, s_{n+1} \in \mathcal{W}_{Mono}$, we get $j_{n+1} \in \mathcal{W}_{Mono}$. Hence $j_{n+1} \in \mathcal{W}_{Mono} \cap Mono$. \square

Theorem 3.4. **SSimp** is a left-determined model category with $\mathcal{C} = \mathbf{Mono}$ and $\mathcal{F} = \mathcal{J}^\square$.

Proof. $(\mathbf{Mono}, \mathbf{Mono}^\square)$ is a weak factorization system (see [B] or [AHRT2]); analogously for $(\mathrm{cof}(\mathcal{J}), \mathcal{J}^\square)$. To prove the result it suffices to show that

$$\mathcal{W}_{\mathbf{Mono}} \cap \mathbf{Mono} = \mathrm{cof}(\mathcal{J})$$

(cf. [B]). Following Lemma 3.3, we have

$$\mathrm{cof}(\mathcal{J}) \subseteq \mathcal{W}_{\mathbf{Mono}} \cap \mathbf{Mono}.$$

The opposite inclusion will follow from properties of the adjunction $L \dashv H$ between symmetric simplicial sets and simplicial sets.

Since L preserves monomorphisms, H preserves trivial fibrations, i.e.,

$$H(\mathbf{Mono}^\square) \subseteq \mathcal{W},$$

where \mathcal{W} denotes the class of weak equivalences of simplicial sets. Since H preserves monomorphisms as well, we have

$$H(\mathcal{W}_{\mathbf{Mono}} \cap \mathbf{Mono}) \subseteq \mathcal{W} \cap \mathbf{Mono}^* = \mathrm{cof}(\mathcal{J}^*)$$

where \mathbf{Mono}^* denotes the monomorphisms in **Simp** and \mathcal{J}^* is the generating set of horns in simplicial sets (cf. [Ho]). Since $L(\mathcal{J}^*) = \mathcal{J}$, we have $L(\mathrm{cof}(\mathcal{J}^*)) \subseteq \mathrm{cof}(\mathcal{J})$. Consequently,

$$LH(\mathcal{W}_{\mathbf{Mono}} \cap \mathbf{Mono}) \subseteq \mathrm{cof}(\mathcal{J}).$$

The functor H sends a symmetric simplicial set $A = (UA, \mathcal{A})$ to the simplicial set HA having as (oriented) simplices all possible orientations of simplices from \mathcal{A} . The functor L then produces from each orientation a non-oriented simplex in LHA . Hence LH multiplies each non-degenerated n -simplex in A $n!$ times. By sending each simplex in \mathcal{A} to its standard orientation, we get a natural transformation $\varrho : Id \rightarrow LH$ that splits the adjunction counit $\varepsilon : LH \rightarrow Id$. Hence each morphism f in **SSimp** is a retract of $LH(f)$. Consequently,

$$\mathcal{W}_{\mathbf{Mono}} \cap \mathbf{Mono} \subseteq \mathrm{cof}(\mathcal{J}).$$

□

Remark 3.5. Both L and H are left Quillen functors. Moreover, $L \dashv H$ is a Quillen equivalence. Following [Ho, 1.3.13], this amounts to showing that

$$X \xrightarrow{\eta_X} HLX \xrightarrow{Hr_{LX}} H(LX)_f$$

is a weak equivalence for each simplicial set X (where η is the adjunction unit and $r_{LX} : LX \rightarrow (LX)_f$ is a fibrant replacement) and that

$$\varepsilon_Y : LHY \rightarrow Y$$

is a weak equivalence for each fibrant symmetric simplicial set Y . But η_X is a trivial cofibration because it is given by completing horns to simplices, and Hr_{LX} is a trivial cofibration too, because H is a left Quillen functor. That ε_Y is a trivial fibration follows from its description given in the proof above.

As a consequence we obtain that **Simp** and **SSimp** have equivalent homotopy categories.

Remark 3.6. The model category **Simp** is not left-determined. To prove this we consider the class \mathcal{X} of morphisms $f : A \rightarrow B$ such that one of the following possibilities occurs (U^* denotes the underlying functor **Simp** \rightarrow **Set**):

- (a) there are vertices $b_1 \in U^*B$, $b_2 \in U^*B - (U^*f)(U^*A)$, an edge e in B from b_1 to b_2 , but no edge in B from b_2 to b_1 ;
- (b) there are vertices $a_1, a_2 \in U^*A$ and an edge e in A from a_1 to a_2 such that there is no edge in A from a_2 to a_1 , but there is an edge in B from $U^*(f)(a_2)$ to $U^*(f)(a_1)$;
- (c) there are vertices $a_1, a_2 \in U^*A$ and an edge e in B from $U^*(f)(a_1)$ to $U^*(f)(a_2)$, but there is no edge in B from $U^*(f)(a_2)$ to $U^*(f)(a_1)$ and no edge in A from a_1 to a_2 .

Since (the oriented) horn $j_1^* : \Delta_1^0 \rightarrow \Delta_1$ belongs to \mathcal{X} , it suffices to show that $\mathcal{X} \cap \mathcal{W}_{Mono^*} = \emptyset$. But $\mathcal{X} \cap Mono^\square = \emptyset$ and no element of \mathcal{X} can arise by operations 2.1 (ii) and (iii) from morphisms not belonging to \mathcal{X} .

Remark 3.7. **SComp** is a left-determined model category with $\mathcal{C} = Mono$ and $\mathcal{F} = \mathcal{J}^\square$. In fact, both i_n , $n \geq 0$ and j_n , $n > 0$ are morphisms of simplicial complexes. Hence the result follows from Theorem 3.4.

Corollary 3.8. *For each small category \mathcal{X} the functor category $\mathbf{SSimp}^\mathcal{X}$ is a left-determined model category (with the Bousfield-Kan model category structure).*

The proof follows from Theorem 3.4 and Proposition 2.5.

4. UNIVERSAL MODEL CATEGORIES

We will show that $\mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$ is a universal model category over \mathcal{X} in the sense of D. Dugger [D] for each small category \mathcal{X} . In particular, **SSimp** is a universal model category over the one-morphism category. We will denote by

$$Y^* : \mathcal{X} \longrightarrow \mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$$

the composition

$$\mathcal{X} \xrightarrow{Y_\mathcal{X}} \mathbf{Set}^{\mathcal{X}^{\text{op}}} \xrightarrow{D_\mathcal{X}} (\mathbf{Set}^{\mathcal{X}^{\text{op}}})^{\mathbf{F}^{\text{op}}}$$

where $D_\mathcal{X}$ is a left adjoint to the underlying functor

$$U_\mathcal{X} : (\mathbf{Set}^{\mathcal{X}^{\text{op}}})^{\mathbf{F}^{\text{op}}} \longrightarrow \mathbf{Set}^{\mathcal{X}^{\text{op}}}$$

given by evaluation at 1, i.e., $U_\mathcal{X} = ev_1$. Of course, we use the identifications

$$(\mathbf{Set}^{\mathbf{F}^{\text{op}}})^{\mathcal{X}^{\text{op}}} \cong \mathbf{Set}^{(\mathbf{F} \times \mathcal{X})^{\text{op}}} \cong (\mathbf{Set}^{\mathcal{X}^{\text{op}}})^{\mathbf{F}^{\text{op}}}.$$

Objects of $\mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$ may be called symmetric simplicial presheaves; then $D_\mathcal{X}(A)$ is the discrete symmetric simplicial presheaf over A . We also have the Yoneda embedding

$$\overline{Y} : \mathbf{F} \times \mathcal{X} \longrightarrow \mathbf{SSimp}^{\mathcal{X}^{\text{op}}},$$

and we will use the notation

$$\Delta_{n,X} = \overline{Y}(n+1, X)$$

for $n \geq 0$ and $X \in \mathcal{X}$. It is easy to see that

$$\Delta_{n,X} = F_X(\Delta_n)$$

where $F_X : \mathbf{SSimp} \longrightarrow \mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$ is a left adjoint to the evaluation functor ev_X . We will also denote

$$\partial \Delta_{n,X} = F_X(\partial \Delta_n).$$

Theorem 4.1. *Let \mathcal{K} be a functorial model category, \mathcal{X} a small category and $H : \mathcal{X} \rightarrow \mathcal{K}$ a functor such that all objects HX , $X \in \mathcal{X}$ are cofibrant. Then there is a left Quillen functor $H^* : \mathbf{SSimp}^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$ such that $H^* \cdot Y^* = H$.*

Proof. Let \mathbf{F}_n be the full subcategory of \mathbf{F} consisting of cardinals $0 < k \leq n+1$. We get the induced inclusions

$$\mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}} \subseteq \mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$$

where $\mathbf{SSimp}_n = \mathbf{Set}^{\mathbf{F}_n^{\text{op}}} (\mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}} \hookrightarrow \mathbf{SSimp}^{\mathcal{X}^{\text{op}}}$ is given by $\mathbf{SSimp}_n \hookrightarrow \mathbf{SSimp}$, which is induced by the functor $\mathbf{F}_n \hookrightarrow \mathbf{F} \xrightarrow{Y} \mathbf{SSimp}$). In particular, $\mathbf{SSimp}_0^{\mathcal{X}^{\text{op}}} \cong \mathbf{Set}^{\mathcal{X}^{\text{op}}}$ is the category of discrete symmetric simplicial presheaves. Since \mathcal{K} is cocomplete and $\mathbf{Set}^{\mathcal{X}^{\text{op}}}$ is a free cocompletion of \mathcal{X} (see [AR, 1.45]), H extends to a colimit preserving functor

$$H_0^* : \mathbf{SSimp}_0^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$$

such that $H_0^* Y = H$.

Assume that we have the functor

$$H_n^* : \mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$$

extending H_{n-1}^* . Since $\partial\Delta_{n+1,X}$, $\Delta_{0,X}$ belong to $\mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}}$ for $X \in \mathcal{X}$, we can define $H_{n+1}^*(\Delta_{n+1,X})$ by the functorial (cofibration, trivial fibration) factorization

$$H_n^*(\partial\Delta_{n+1,X}) \xrightarrow{c_{n+1,X}} H_{n+1}^*(\Delta_{n+1,X}) \xrightarrow{r_{n+1,X}} H_n^*(\Delta_{0,X})$$

of the morphism $H_n^*(F_X(p_n)) : H_n^*(\partial\Delta_{n+1,X}) \rightarrow H_n^*(\Delta_{0,X})$ where $p_{n+1} : \partial\Delta_{n+1} \rightarrow \Delta_0$. To get an extension H_{n+1}^* of H_n^* , below we will define

- (a) $H_{n+1}^*(f)$ for $f = \bar{Y}(f_1, f_2) : \Delta_{n+1,X} \rightarrow \Delta_{n+1,Y}$ where $f_1 : \{0, 1, \dots, n+1\} \rightarrow \{0, 1, \dots, n+1\}$ is a bijection,
- (b) $H_{n+1}^*(t)$ for $t = \bar{Y}(t_1, t_2) : \Delta_{m,X} \rightarrow \Delta_{n+1,Y}$ where $m \leq n$,

and

- (c) H_{n+1}^* for $u = \bar{Y}(u_1, u_2) : \Delta_{n+1,X} \rightarrow \Delta_{m,Y}$ where $m \leq n$.

(a) f_1 induces the isomorphisms $\bar{f}_1 : \Delta_{n+1} \rightarrow \Delta_{n+1}$, $\partial\bar{f}_1 : \partial\Delta_{n+1} \rightarrow \partial\Delta_{n+1}$ and f_2 induces a natural transformation $\varphi_{f_2} : F_X \rightarrow F_Y$. Hence f induces the homomorphism $\partial f : \partial\Delta_{n+1,X} \rightarrow \partial\Delta_{n+1,Y}$. We define $H_{n+1}^*(f)$ by the functorial filling

$$\begin{array}{ccccc} H_n^*(\partial\Delta_{n+1,X}) & \xrightarrow{c_{n+1,X}} & H_{n+1}^*(\Delta_{n+1,X}) & \xrightarrow{r_{n+1,X}} & H_n^*(\Delta_{0,X}) \\ \downarrow H_n^*(\partial f) & & \downarrow H_{n+1}^*(f) & & \downarrow H_n^*((\varphi_{f_2})_{\Delta_0}) \\ H_n^*(\partial\Delta_{n+1,Y}) & \xrightarrow{c_{n+1,Y}} & H_{n+1}^*(\Delta_{n+1,Y}) & \xrightarrow{r_{n+1,Y}} & H_n^*(\Delta_{0,Y}) \end{array}$$

- (b) Since t factorizes through $i_{n+1,Y}$, i.e.,

$$t : \Delta_{m,X} \xrightarrow{t'} \partial\Delta_{n+1,Y} \xrightarrow{i_{n+1,Y}} \Delta_{n+1,Y},$$

we put $H_{n+1}^*(t) = c_{n+1,Y} \cdot H_n^*(t')$.

(c) To define $H_{n+1}^*(u)$ for each u , it suffices to do this for the retraction u^0 of $\Delta_{n+1,X}$ to one of its faces $\Delta_{n,X}$. In this case, we take $H_{n+1}^*(u^0)$ given by the lifting property

$$\begin{array}{ccc}
 H_n^*(\partial\Delta_{n+1,X}) & \xrightarrow{c_{n+1,X}} & H_{n+1}^*(\Delta_{n+1,X}) \\
 \downarrow H_n^*(u^0 \cdot i_{n+1,X}) & \nearrow H_{n+1}^*(u^0) & \downarrow r_{n+1,X} \\
 H_n^*(\Delta_{n,X}) & \xrightarrow{H_n^*(s_{n,X})} & H_n^*(\Delta_{0,X})
 \end{array}$$

where $s_{n,X} = \overline{Y}(s, \text{id}_X)$ and $s : n+1 \rightarrow 1$.

To prove that H_{n+1}^* is a functor, it suffices to consider the following cases:

(1) In

$$\begin{array}{ccccc}
 & & H_n^*(\Delta_{m,X}) & & \\
 & \swarrow H_n^*(t') & \downarrow H_{n+1}^*(t) & & \\
 H_n^*(\partial\Delta_{n+1,Y}) & \xrightarrow{c_{n+1,Y}} & H_{n+1}^*(\Delta_{n+1,Y}) & \xrightarrow{r_{n+1,Y}} & H_n^*(\Delta_{0,Y}) \\
 \downarrow H_n^*(\partial f) & & \downarrow H_{n+1}^*(f) & & \downarrow H_n^*((\varphi_{f_2})_{\Delta_0}) \\
 H_n^*(\partial\Delta_{n+1,Z}) & \xrightarrow{c_{n+1,Z}} & H_{n+1}^*(\Delta_{n+1,Z}) & \xrightarrow{r_{n+1,Z}} & H_n^*(\Delta_{0,Z})
 \end{array}$$

we have $(ft)' = \partial f \cdot t'$ and thus $H_{n+1}^*(f) \cdot H_{n+1}^*(t) = H_{n+1}^*(ft)$.

(2) In

$$\begin{array}{ccc}
 & & H_n^*(\Delta_{m,X}) \\
 & \swarrow H_n^*(t') & \downarrow H_{n+1}^*(t) \\
 H_n^*(\partial\Delta_{n+1,Y}) & \xrightarrow{c_{n+1,Y}} & H_{n+1}^*(\Delta_{n+1,Y}) \\
 \downarrow H_n^*(u^0 \cdot i_{n+1,Y}) & \nearrow H_{n+1}^*(u^0) & \downarrow r_{n+1,Y} \\
 H_n^*(\Delta_{n,Y}) & \xrightarrow{H_n^*(s_{n,Y})} & H_n^*(\Delta_{0,Y})
 \end{array}$$

we have $u^0 \cdot t = u^0 \cdot i_{n+1,Y} \cdot t'$ and thus $H_{n+1}^*(u^0) \cdot H_{n+1}^*(t) = H_{n+1}^*(u^0 \cdot t)$.

We have defined H_{n+1}^* on the image of the Yoneda embedding

$$\bar{Y}_n : \mathbf{F}_n \times \mathcal{X} \rightarrow \mathbf{Set}^{(\mathbf{F}_n \times \mathcal{X})^{\text{op}}} \cong \mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}}.$$

Since this is a free cocompletion of $\mathbf{F}_n \times \mathcal{X}$, we obtain a colimit-preserving functor $H_{n+1}^* : \mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$ extending H_n^* .

We have constructed an increasing chain of colimit-preserving functors $H_n^* : \mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$ for $n = 0, 1, \dots$, which yields a colimit-preserving functor

$$H^* : \mathbf{SSimp}^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$$

with the restriction H_n^* on $\mathbf{SSimp}_n^{\mathcal{X}^{\text{op}}}$. Moreover, H^* is a left adjoint functor (see (a) in the proof of 1.45 in [AR]). In particular, $H^*Y^* = H_0^*Y^* = H$. It remains to be proved that H^* preserves cofibrations and trivial cofibrations. Since $\partial\Delta_{n+1,X}$ is a colimit of $\partial_{m,X}$, $m \leq n$, by (b) of the construction we get that $H^*(i_{n+1,X}) = c_{n+1,X}$. Hence H^* preserves cofibrations (using Lemma 3.2 and the fact that H^* preserves colimits). Since $s_{n+1,X} = s_{n,X} \cdot u^0$, by (c) of the construction we get that $H^*(s_{n+1,X}) = r_{n+1,X}$. Hence $H^*(s_{n+1,X})$ is a weak equivalence in \mathcal{K} for $n > 0$ and $X \in \mathcal{X}$. Since $s_{m,X} \cdot f = s_{n,Y}$ for each morphism $f : \Delta_{n,X} \rightarrow \Delta_{m,Y}$, $H^*(f)$ is a weak equivalence in \mathcal{K} as well. In particular, $H^*F_X(j_1) : H^*F_X(\Delta_1^0) = H^*\Delta_{0,X} \rightarrow H^*\Delta_{1,X}$ is a weak equivalence. Since it is also a cofibration, it is a trivial cofibration. Assume that $H^*F_X(j_1), \dots, H^*F_X(j_n)$ are trivial cofibrations in \mathcal{K} . In the notation of Lemma 3.3, we get that $H^*F_X(g_n)$ is a trivial cofibration in \mathcal{K} because $u_1^0 = j_1$. Since by assumption $H^*F_X(j_1), \dots, H^*F_X(j_n)$ are trivial cofibrations, $H^*F_X(q_n)$ is a trivial cofibration too. Therefore $H^*F_X(t_n)$ is a weak equivalence in \mathcal{K} and thus $H^*F_X(j_{n+1})$ is a weak equivalence too. Since it is a cofibration, it is a trivial cofibration. \square

Remark 4.2. Let \mathcal{K} be a functorial model category, \mathcal{X} a small category and $H : \mathcal{X} \rightarrow \mathcal{K}$ an arbitrary functor. Then there is a left Quillen functor $H^* : \mathbf{SSimp}^{\mathcal{X}^{\text{op}}} \rightarrow \mathcal{K}$ and a natural transformation $\gamma : H^*Y^* \rightarrow H$ which is a pointwise trivial fibration in \mathcal{K} .

In fact, let $H_0 : \mathcal{X} \rightarrow \mathcal{K}$ be the composition $Q \cdot H$ where $Q : \mathcal{K} \rightarrow \mathcal{K}$ is the (functorial) cofibrant replacement functor and $\gamma = qH : H_0 \rightarrow H$ the corresponding natural transformation. Then γ is a pointwise trivial fibration in \mathcal{K} (because $q : Q \rightarrow \text{Id}_{\mathcal{K}}$ is). If we start the construction of the functor H^* with H_0 (i.e., $H_0^* \cdot Y^* = H_0$), we get the result.

Corollary 4.3. *Let \mathcal{K} be a functorial model category and K an object in \mathcal{K} . Then there is a left Quillen functor $H^* : \mathbf{SSimp} \rightarrow \mathcal{K}$ and a trivial fibration $\gamma : H^*(\Delta_0) \rightarrow K$.*

This is a special case of Theorem 4.1 (for \mathcal{X} a one-morphism category).

Remark 4.4. Again if K is cofibrant, then $H^*(\Delta_0) = K$.

REFERENCES

- [AHRT1] J. Adámek, H. Herrlich, J. Rosický, and W. Tholen, On a generalized small-object argument for the injective subcategory problem, *Cahiers Topologie Géom. Différentielle Catég.* XLIII (2002), 83-106.
- [AHRT2] J. Adámek, H. Herrlich, J. Rosický, and W. Tholen, Weak factorization systems and topological functors, *Applied Categ. Structures* 10 (2002), 237-249.

- [AR] J. Adámek and J. Rosický, Locally presentable and accessible categories, London Mathematical Society Lecture Notes Series, Vol. 189, Cambridge University Press, Cambridge, 1994. MR **95j**:18001
- [B] T. Beke, Sheafifiable homotopy model categories, Math. Proc. Cambridge Philos. Soc. 129 (2000), 447–475. MR **2001i**:18015
- [CSS] C. Casacuberta, D. Sceveneles, and J. H. Smith, Implications of large-cardinal principles in homotopical localizations, preprint, 1998.
- [D] D. Dugger, Universal homotopy theories, Adv. Math. 164 (2001), 144–176. MR **2002k**:18021
- [G] M. Grandis, Higher fundamental functors for simplicial sets, Cahiers Topologie Géom. Différentielle Catég. 42 (2001), 101–136. MR **2002f**:18026
- [H] P. S. Hirschhorn, Localization of Model Categories, preprint, 1998, <http://www.math.unit.edu/psh>.
- [Ho] M. Hovey, Model Categories, Amer. Math. Soc., Providence, RI, 1999. MR **99h**:55031
- [EZ] J. Eilenberg and J. A. Zilber, Semi-simplicial complexes and singular homology, Ann. of Math. 51 (1950), 499–513. MR **11**:734e
- [L] F. W. Lawvere, Toposes generated by codiscrete objects, in Combinatorial Topology and Functional Analysis, Notes 1988, 1989, 1992.
- [Q] D. Quillen, Homotopical algebra, Lecture Notes in Math. 43, Springer-Verlag, Berlin, 1967. MR **36**:6480
- [RT] J. Rosický and W. Tholen, Lax factorization algebras, J. Pure Appl. Algebra 175 (2002), 355–382.
- [S] J. H. Smith, Combinatorial model categories, in preparation.

DEPARTMENT OF MATHEMATICS, MASARYK UNIVERSITY, 662 95 BRNO, CZECH REPUBLIC
E-mail address: rosicky@math.muni.cz

DEPARTMENT OF MATHEMATICS AND STATISTICS, YORK UNIVERSITY, TORONTO M3J 1P3,
CANADA
E-mail address: tholen@pascal.math.yorku.ca

THE COMBINATORIAL RIGIDITY CONJECTURE IS FALSE FOR CUBIC POLYNOMIALS

CHRISTIAN HENRIKSEN

ABSTRACT. We show that there exist two cubic polynomials with connected Julia sets which are combinatorially equivalent but not topologically conjugate on their Julia sets. This disproves a conjecture by McMullen from 1995.

INTRODUCTION AND RESULT

Let $\mathcal{P}_d = \{z^d + a_{d-2}z^{d-2} + \cdots + a_0\} \leftrightarrow \mathbb{C}^{d-1}$ be the space of monic centered polynomials of degree $d > 1$.

Our object is to show that there exists a cubic polynomial $f \in \mathcal{P}_3$ that is not combinatorially rigid. To specify what this means, we need to introduce some notation and results from the theory of holomorphic dynamics, and, more specifically, the dynamics of polynomials. We will assume that the reader has some knowledge of the theory of iteration of polynomials in one variable. There are several introductory books on holomorphic dynamics, such as [M1]; also Lyubich [L1] has recently published a survey article.

The filled Julia set $K(f)$ of $f \in \mathcal{P}_d$ is the set of values of z such that the orbit $z, f(z), f \circ f(z) = f^2(z), \dots$ is bounded. The boundary of this set $J(f) = \partial K(f)$ is called the Julia set of f , and it coincides with the set of points that admit no neighborhood restricted to which the iterates id, f, f^2, \dots form a normal family.

We denote by \mathcal{C}_d the connectedness locus of the degree d monic centered polynomials:

$$\mathcal{C}_d = \{f \in \mathcal{P}_d \mid K(f) \text{ is connected}\}.$$

The connectedness locus of the quadratic polynomials, \mathcal{C}_2 , is the Mandelbrot set, and we also denoted it by M .

Given $f \in \mathcal{C}_d$, there exists a unique conformal isomorphism

$$\phi_f : \mathbb{C} \setminus K(f) \xrightarrow{\cong} \mathbb{C} \setminus \overline{\mathbb{D}}$$

that conjugates f to $z \mapsto z^d$ and satisfies $\phi_f(z) = z + \mathcal{O}(1/z)$. The conformal isomorphism is called a Böttcher coordinate.

Using the Böttcher coordinate, we can define the notion of the dynamical ray $R_f(\theta)$ of angle $\theta \in \mathbb{R}/\mathbb{Z}$ as the preimage $\phi_f^{-1}\{\exp(\eta + 2i\pi\theta) \mid \eta > 0\}$ of a radial segment. For a polynomial $f \in \mathcal{P}_d$, we define the *potential* $G_f : \mathbb{C} \setminus K(f) \rightarrow \mathbb{R}$ by

$$G_f(z) = \lim_{n \rightarrow +\infty} \frac{1}{d^n} \log |f^n(z)|.$$

Received by the editors January 30, 2002 and, in revised form, August 13, 2002.

2000 *Mathematics Subject Classification*. Primary 37F10; Secondary 37F20, 37F45.

This research was funded by a Marie Curie Fellowship.

Then, when $K(f)$ is connected, $G_f = \log |\phi_f|$, and a dynamical ray is parallel to the vector field $\text{grad } G_f$. This allows us to generalize the notion of a dynamical ray. Indeed, for an arbitrary polynomial $f \in \mathcal{P}_d$, the dynamical ray $R_f(\theta)$ is the curve that is tangent to $\text{grad } G_f$ and tangent to the segment $\{r \exp(2i\pi\theta) \mid r > 0\}$ near infinity.

Following a dynamical ray down potential, two things might happen. Either the ray hits a critical point of G_f and we say it bifurcates, or it accumulates on the Julia set. The first behaviour only takes place when G_f has critical points, which is equivalent to f having critical points that are attracted to infinity and to K_f being disconnected.

Suppose the ray $R_f(\theta)$ does not bifurcate. Consider $\overline{R_f(\theta)} \setminus R_f(\theta)$. This continuum is called the accumulation set of the ray. If it consists of only one point, we say that the ray *lands*. In the connected case, $f \in \mathcal{C}_d$, that means that the limit $z_0 = \lim_{\eta \searrow 0} \phi_f^{-1} \circ \exp(\eta + 2i\pi\theta)$ exists.

Rays of rational angle are special, in that they always land when $f \in \mathcal{C}_d$:

Proposition 1. *Let $f \in \mathcal{P}_d$ and $\theta \in \mathbb{Q}/\mathbb{Z}$. Then the ray $R_f(\theta)$ either bifurcates or lands at a (pre)periodic point γ . In the latter case, γ is periodic if θ is periodic, and preperiodic if θ is preperiodic; it is either repelling or parabolic.* \square

The proposition is proved in the Orsay Notes (Proposition 2 of exposé 8 in [DH1]).

This proposition allows us to distinguish the two fixed points of a quadratic polynomial $P_c : z \mapsto z^2 + c$ when $c \in M \setminus \{1/4\}$ (for $c = 1/4$, P_c has one fixed point and it is of multiplicity 2). Indeed, the ray $R_{P_c}(0)$ is fixed by P_c and then must land at a fixed point. This fixed point is called the β -fixed point, and the other fixed point is called the α -fixed point.

Definition 1. Following McMullen ([Mc1]) we define the *rational lamination* $\lambda_{\mathbb{Q}}(f) \subset \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}$ of $f \in \mathcal{C}_d$ to be the equivalence relation under which two rational angles θ' and θ'' are equivalent if and only if the two dynamical rays $R_f(\theta')$ and $R_f(\theta'')$ land at the same point.

Definition 2. A polynomial $f \in \mathcal{C}_d$ with no indifferent cycles is said to be *combinatorially rigid* if for every $g \in \mathcal{C}_d$ with no indifferent cycles such that $\lambda_{\mathbb{Q}}(f) = \lambda_{\mathbb{Q}}(g)$, the composition of Böttcher coordinates $\phi_g^{-1} \circ \phi_f$ extends to a quasiconformal homeomorphism on the whole Riemann sphere.

Note that we do not require the extension to be a conjugacy on the sphere; however, by continuity, it will be a conjugacy on the Julia set $J(f)$.

The following statement has been conjectured:

Every polynomial $f \in \mathcal{C}_d$ without indifferent periodic cycles is combinatorially rigid. (See [Mc1].)

This conjecture can be viewed as a series of conjectures, one for each value of $d = 2, 3, \dots$

If the conjecture is true for quadratic polynomials ($d = 2$), then it would imply that the Mandelbrot set M is locally connected [Sch]. Local connectivity of M is one of the most important questions in the study of the dynamics of quadratic polynomials. It has been verified for most types of parameters in the boundary of M . A summary of known results on this question is given by Lyubich in [L2], Appendix B.

We shall show that the conjecture is false in degree 3, by finding two cubic polynomials with connected Julia sets, with no indifferent cycles and the same rational lamination, but which fail to be even topologically conjugate on their Julia sets.

A sketch of the proof follows: From work of Sørensen [Sø] it follows that there is an infinitely renormalizable quadratic polynomial $P_{c_\infty} : z \mapsto z^2 + c_\infty$ whose Julia set contains a subcontinuum without periodic or preperiodic points. Choose a quadratic polynomial Q whose critical point eventually lands at the β -fixed point. Let f_0 be the result of intertwining Q with P_{c_∞} at their β -fixed points, using the intertwining surgery construction of Epstein and Yampolsky [EY].

The cubic polynomial f_0 has a quadratic-like restriction $f_0 : U' \rightarrow U$ that is hybrid equivalent to P_{c_∞} .

There is a neighborhood Λ of f_0 such that for each f_λ there are disks U'_λ, U_λ for which the restriction $f_\lambda : U'_\lambda \rightarrow U_\lambda$ is quadratic-like. This can be accomplished so that the family $(f_\lambda : U'_\lambda \rightarrow U_\lambda)$ is an *analytic family* in the sense of [DH2]. Using results in [DH2], we can suppose that the set of parameters for which the corresponding quadratic-like mappings all are hybrid equivalent to P_{c_∞} locally forms a codimension-one submanifold of the space of cubic polynomials.

For parameters t in this submanifold, one critical point $\omega_1(t)$ belongs to this quadratic-like restriction, whereas the other $\omega_2(t)$ does not. Using holomorphic motion arguments, we see that there are copies of (part of) the Julia set of P_{c_∞} showing up in the set of t -values for which f_t has connected Julia set. Loosely speaking, the position of the parameter t in such a parameter copy corresponds to the position of an iterate of $\omega_2(t)$ in a copy of $K(P_{c_\infty})$, in the dynamical plane. So by varying t we can slide an iterate of $\omega_2(t)$ around in a continuum without periodic or preperiodic points. It follows that no rational ray bifurcates. However, we shall see that the dynamics on the Julia sets do change.

1. A SPECIAL QUADRATIC POLYNOMIAL

In this section we will find a quadratic polynomial that does not have indifferent cycles and whose Julia set contains a continuum with no periodic or preperiodic cycles.

Polynomial-like mappings and renormalization.

Definition 3. If U', U are topological disks with U' compactly contained in U and $f : U' \rightarrow U$ is a holomorphic, ramified covering of degree $d > 1$, then $f : U' \rightarrow U$ is called a *polynomial-like map* of degree d .

The notion of polynomial-like mappings was introduced by Douady and Hubbard (see [DH2]) and is extremely useful. When the degree of a polynomial-like map is two, we call it *quadratic-like*. As the name suggests, a quadratic-like map behaves qualitatively like a quadratic polynomial.

Analogously to the definitions for polynomials, we define the *filled Julia set* of a polynomial-like map f as the set of points that do not escape under iteration: $K(f) = \{z \in U' \mid f^1(z) \in U', f^2(z) \in U', \dots\}$, and the *Julia set* $J(f)$ of f as its boundary: $J(f) = \partial K(f)$. Two polynomial-like mappings f, g are called *hybrid equivalent* if there exists a quasiconformal mapping ϕ that maps a neighborhood of $K(f)$ onto a neighborhood of $K(g)$, conjugating f to g and satisfying $\frac{\partial \phi}{\partial \bar{\phi}} = \frac{\partial \phi / \partial \bar{z}}{\partial \phi / \partial z} =$

0 almost everywhere on $K(f)$. This clearly defines an equivalence relation, whose equivalence classes are called *hybrid classes*.

That quadratic-like mappings behave like quadratic polynomials follows from Douady and Hubbard's Straightening Theorem, which for quadratic-like mappings boils down to:

Straightening Theorem. *A quadratic-like map f is hybrid equivalent to a quadratic polynomial. If $K(f)$ is connected, f is hybrid equivalent to a unique monic centered polynomial $z \mapsto z^2 + \chi(f)$.* \square

It may happen that an iterate of a quadratic polynomial in some region behaves like a quadratic-like map.

Definition 4. A quadratic polynomial $P : z \mapsto z^2 + c$ is called *n-renormalizable* for an integer $n > 1$, if there exists a restriction $P^n : U' \rightarrow U$ that is quadratic-like with connected Julia set and $0 \in U'$. The restriction $P^n : U' \rightarrow U$ is called a *renormalization* of P^n , and the set of positive integers $n > 1$ such that P is *n-renormalizable* is called the *levels of renormalization* of P and is denoted $\mathcal{R}(P)$. If this set is nonempty, P is called *renormalizable*, and if it is infinite, P is called *infinitely renormalizable*.

In [Mc2] McMullen shows:

Theorem 2 (Uniqueness of renormalization). *Any two renormalizations of P^n have the same filled Julia set.* \square

An infinitely renormalizable quadratic polynomial. If $n \in \mathcal{R}(P)$, we denote by $K_n(P)$ the uniquely determined filled Julia set of $P^n : U' \rightarrow U$, and we set $J_n(P) = \partial K_n(P)$.

In [Sø], Sørensen proves the following result:

Theorem 3. *There is an infinitely renormalizable quadratic polynomial $P_{c_\infty} : z \mapsto z^2 + c_\infty$ with the following properties: There is a sequence of integers $n_1 < n_2 < \dots$ such that each $n_k \in \mathcal{R}(P_{c_\infty})$, $K_{n_k} \supset K_{n_{k+1}}$ and the continuum $J_\infty = \bigcap K_{n_k}(P_{c_\infty})$ is nontrivial (i.e., contains more than one point).* \square

A proof can also be found in [M2].

Let P_{c_∞} and J_∞ be as given by Theorem 3. Since J_∞ is obtained as an intersection of nested full continua, J_∞ is a full continuum.

By the following theorem of McMullen, J_∞ contains no periodic or preperiodic points.

Theorem 4. *If a quadratic polynomial P is infinitely renormalizable, then:*

- (1) *all periodic cycles of P are repelling;*
- (2) *the filled Julia set $K(P)$ has no interior; and*
- (3) *if $R \subset \mathcal{R}(P)$ is infinite, then the continuum $\bigcap_{n \in R} K_n(P)$ contains no periodic or preperiodic points.*

Proof. This is an immediate consequence of [Mc2], Theorems 8.1 and 7.8. \square

Arbitrarily small copies of J_∞ accumulate on the β -fixed point of P_{c_∞} :

Proposition 5. *Given any $\epsilon > 0$, there are an integer $n > 0$ and a topological disk contained in $\{|z - \beta| < \epsilon\}$ that $P_{c_\infty}^n$ maps univalently onto a neighborhood of J_∞ .*

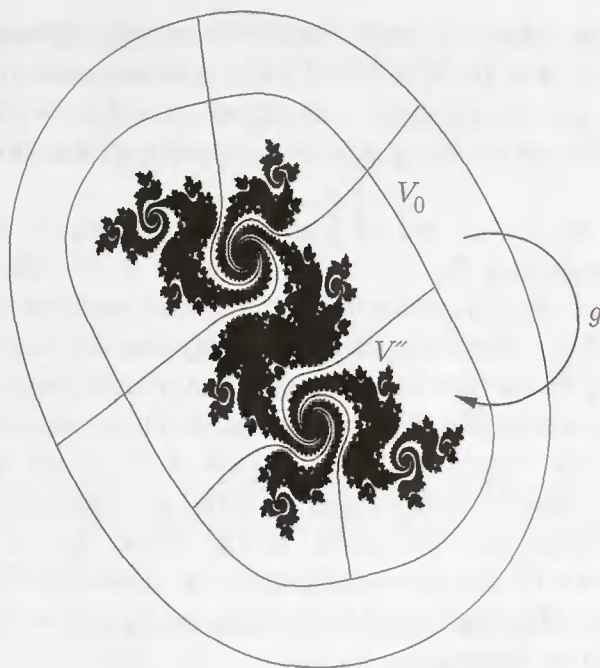


FIGURE 1. The inverse branch g used in the proof of Proposition 5. Due to inherent difficulties in drawing the filled Julia set $K(P_{c_{\infty}})$, a filled Julia set for another parameter is drawn.

Proof. By Theorem 4, the α -fixed point, α , of $P_{c_{\infty}}$ is repelling, since $P_{c_{\infty}}$ is infinitely renormalizable. By a result of Douady and Yoccoz, k rational rays land at α for some $k > 0$. The ray of angle 0° lands at the β -fixed point, β , and it follows that $k > 1$. The region $W = \{z \mid G_{P_{c_{\infty}}} < 1\}$ is a Jordan domain. The k rational rays landing at α cut this region into k pieces V_0, V_1, \dots, V_{k-1} , which we enumerate cyclically with respect to α in such a way that the critical point 0 is contained in V_0 . The preimage W' of W is cut into $2k - 1$ regions $V'_0, V'_1, \dots, V'_{2k-2}$ by the rational rays landing at α and $-\alpha$. We can assume $0 \in V'_0$. Each piece $V'_i, i \neq 0$, is mapped properly of degree one to a piece $V_{\sigma(i)}$, and V'_0 is mapped properly of degree 2 to a piece $V_{\sigma(0)}$. Since none of the rational rays landing at α are fixed, $V_{\sigma(0)}$ does not coincide with V_0 . It follows that V_0 has two preimages: one V'' that is compactly contained in V_0 , and the symmetric one $-V''$. Since 0 is contained in J_{∞} and J_{∞} contains neither α nor $-\alpha$, we get that J_{∞} is contained in V_0 and does not intersect V'' .

Now consider the inverse branch $g = P_{c_{\infty}}^{-1} : V_0 \rightarrow V''$ (see Figure 1). This univalent map is a strong contraction with respect to the hyperbolic metric on V_0 and necessarily contains a fixed point, which then must be β . Now $V_0, g(V_0), g^2(V_0), \dots$ is a sequence of topological disks, each containing β in its closure, whose diameters tend towards zero. Finally, $P_{c_{\infty}}^n$ maps $g^n(V_0)$ univalently onto V_0 . \square

2. CUBIC POLYNOMIALS

In this section we consider the space of cubic polynomials, and see that deformed copies of (parts of) quadratic Julia sets show up both in dynamical planes and in parameter slices.

We parametrize the space of cubic monic centered polynomials by \mathbb{C}^2 , letting $f_{\mathbf{a}}(z) = z^3 - 3a^2z + b$, $\mathbf{a} = (a, b) \in \mathbb{C}^2$. This is a four-to-one covering of the cubic polynomials modulo affine conjugacy, ramified over $\{ab = 0\}$. The two critical points $\omega_1 = \omega_1(\mathbf{a}) = a$ and $\omega_2 = \omega_2(\mathbf{a}) = -a$ are then global holomorphic functions of the parameter \mathbf{a} .

Define $\mathcal{Q}_1 \subset \mathbb{C}^2$ to be the set of parameters for which $f_{\mathbf{a}}$ can be restricted to a quadratic-like mapping $f_{\mathbf{a}} : U' \rightarrow U$ with $\omega_1 \in U'$. We say that \mathcal{Q}_1 is the set of parameters such that $f_{\mathbf{a}}$ is ω_1 -renormalizable, and we call $f_{\mathbf{a}} : U' \rightarrow U$ an ω_1 -renormalization of $f_{\mathbf{a}}$. Similarly, denote by \mathcal{Q}_2 the set of parameters for which $f_{\mathbf{a}}$ is ω_2 -renormalizable. Notice that a cubic polynomial being ω_i -renormalizable is quite different than a quadratic polynomial being renormalizable, since in the former case we are only considering restrictions of the cubic map, whereas in the latter case we only consider restrictions of some n th iterate, $n > 1$. (Also, this terminology is not completely standard, in the sense that it is usually required that the filled Julia set of the renormalization be connected for a mapping to be called renormalizable. However, in this context we find it convenient to drop that requirement for ω_i -renormalizations.)

An ω_1 -renormalization (or an ω_2 -renormalization) of $f_{\mathbf{a}}$ has the following properties, independent of the choice of domains:

Proposition 6. *Suppose $f_{\mathbf{a}_0}$ is ω_1 -renormalizable, and let $f_{\mathbf{a}_0} : U'_1 \rightarrow U_1$ and $f_{\mathbf{a}_0} : U'_2 \rightarrow U_2$ denote two ω_1 -renormalizations. Then $K(f_{\mathbf{a}_0} : U'_1 \rightarrow U_1) = K(f_{\mathbf{a}_0} : U'_2 \rightarrow U_2)$. Furthermore, if $K(f_{\mathbf{a}_0} : U'_1 \rightarrow U_1)$ is connected, then both renormalizations are hybrid equivalent to the same quadratic polynomial $z \mapsto z^2 + \chi_1(\mathbf{a}_0)$.*

Proof. The result follows readily from [Mc2] (Theorem 5.11) and from the Straightening Theorem. \square

A one-dimensional submanifold of the cubic polynomials. Let $\mathcal{K}_i \subset \mathcal{Q}_i$, $i = 1, 2$, denote the subset of parameters \mathbf{a} such that an ω_i -renormalization of $f_{\mathbf{a}}$ has connected Julia set.

It follows from Proposition 6 that for a given point $\mathbf{a} \in \mathcal{K}_i$ there is a unique monic centered polynomial $z \mapsto z^2 + \chi_i(\mathbf{a})$ that is hybrid-equivalent to an ω_i -renormalization of $f_{\mathbf{a}}$. This determines mappings χ_i of \mathcal{K}_i into the Mandelbrot set M .

Theorem 7. *There exists a parameter $\mathbf{a}_0 \in \mathcal{K}_1$ satisfying:*

- $\chi_1(\mathbf{a}_0) = c_{\infty}$ (where c_{∞} is the parameter whose existence is guaranteed by Theorem 3);
- an iterate $f_{\mathbf{a}_0}^k(\omega_2(\mathbf{a}_0))$ is equal to the fixed point of $f_{\mathbf{a}_0}$ corresponding to the β -fixed point of $z \mapsto z^2 + c_{\infty}$;
- the set $\chi_1^{-1}(c_{\infty})$ is a one-dimensional submanifold of \mathcal{Q}_1 locally at \mathbf{a}_0 .

The first step in proving Theorem 7 is to establish that $\chi_1^{-1}(c_{\infty})$ is an analytic set in \mathcal{Q}_1 . The definition of an analytic set is as follows.

Definition 5. Suppose D is an open set in \mathbb{C}^n . The set A is called *analytic* in D if for every $z \in D$ there exist an open neighborhood U of z in D and a finite number of holomorphic functions $f_i : U \rightarrow \mathbb{C}$, $i = 1, \dots, k$, such that $A \cap U = \{z \in U \mid f_1(z) = \dots = f_k(z) = 0\}$.

We will give a few properties of analytic sets. Proofs can be found in books on several complex variables, such as [KK] and [GF].

Let $\mathbf{a}_0 \in A$. If there are functions f_1, \dots, f_k , defined on a neighborhood U of \mathbf{a}_0 such that $A \cap U = \{z \in U \mid f_1(z) = \dots = f_k(z) = 0\}$ and

$$\text{rank} \left(\frac{\partial f_i}{\partial z_j} \right) = k,$$

then \mathbf{a}_0 is called a *regular* point or a manifold point. By the Implicit Function Theorem, A is locally biholomorphically equivalent to a domain in \mathbb{C}^{n-k} at \mathbf{a}_0 . We define the dimension of A at \mathbf{a}_0 to be $\dim_{\mathbf{a}_0} A = n - k$. A point in A that is not regular is called *singular*, and we denote the singular points of A by $S(A)$.

Proposition 8. *Let A be an analytic set in the open set $U \subset \mathbb{C}^n$. Then the set of singular points $S(A)$ is analytic in U , and the set of regular points $A \setminus S(A)$ is dense in A .* \square

Clearly, every point is regular and of the same dimension in a neighborhood of a regular point. By the proposition, every singular point lies in the closure of the regular points. The following definition therefore makes sense:

Definition 6. The dimension of an analytic set $A \subset \mathbb{C}^n$, $A \neq \emptyset$, at \mathbf{a}_0 is

$$\limsup_{\mathbf{a} \rightarrow \mathbf{a}_0} \dim_{\mathbf{a}}(A),$$

where \mathbf{a} tends to \mathbf{a}_0 through regular points. We define the dimension of A by

$$\dim A = \sup_{\mathbf{a} \in A} \dim_{\mathbf{a}} A.$$

If $d = \dim A = \inf_{\mathbf{a} \in A} \dim_{\mathbf{a}} A$, we say that A is *purely d -dimensional*.

Theorem 9. *Suppose $A \neq \emptyset$ is analytic in the open set $U \subset \mathbb{C}^n$ and $S(A) \neq \emptyset$. Then*

$$\dim S(A) \leq \dim A - 1.$$

An analytic set is called *irreducible* if it cannot be written as a union of two analytic sets $A = A_1 \cup A_2$ with both A_1 and A_2 proper subsets of A . Every analytic set can be decomposed into irreducible components:

Theorem 10. *Suppose A is analytic in $U \subset \mathbb{C}^n$. Then A has a unique representation $A = \bigcup_{\alpha} A_{\alpha}$ as a locally finite union of irreducible subsets, with $A_{\alpha} \not\subset A_{\beta}$ for $\alpha \neq \beta$. This decomposition is given by*

$$\{A_{\alpha}\}_{\alpha} = \{\overline{C} \mid C \text{ is a connected component of } A \setminus S(A)\}.$$

Here the closure \overline{C} is taken in U . \square

Remark 1. It follows directly from Theorem 10 that if A is analytic in $U \subset \mathbb{C}$ and of dimension 0, then A is discrete in U .

Most of the work to see that $\chi_1^{-1}(c_{\infty})$ is analytic in \mathcal{Q}_1 has been carried out in [DH2]; in fact, they prove (Corollary 2, page 313) that for any analytic family $(f_{\lambda})_{\lambda \in \Lambda}$ of quadratic-like mappings, and any parameter c in the Mandelbrot set, the set of parameters λ for which f_{λ} is hybrid equivalent to $z^2 + c$ is an analytic subset of Λ . The definition of an analytic family of polynomial-like maps is the following.

Definition 7. Let Λ denote a complex analytic manifold and suppose $\mathbf{f} = (f_\lambda : U'_\lambda \rightarrow U_\lambda)_{\lambda \in \Lambda}$ is a family of polynomial-like mappings. Set $\mathcal{U}' = \{(\lambda, z) \mid \lambda \in \Lambda, z \in U'_\lambda\}$, $\mathcal{U} = \{(\lambda, z) \mid \lambda \in \Lambda, z \in U_\lambda\}$ and $f(\lambda, z) = f_\lambda(z)$. We say that \mathbf{f} is an *analytic family of polynomial-like maps* if the following three conditions are satisfied:

- (1) \mathcal{U} and \mathcal{U}' are homeomorphic over Λ to $\Lambda \times \mathbb{D}$;
- (2) the projection π_Λ from the closure of \mathcal{U}' in \mathcal{U} to Λ is proper; and
- (3) $f : \mathcal{U}' \rightarrow \mathcal{U}$ is complex analytic and proper.

For parameters $\mathbf{a} \in \mathcal{Q}_1$, we have that ω_1 -renormalizations of $f_{\mathbf{a}}$ locally form an analytic family. More precisely:

Lemma 1. *Given $\mathbf{a}_0 \in \mathcal{Q}_1$, there are a bi-disk $\Lambda \subset \mathcal{Q}_1$ containing \mathbf{a}_0 and ω_1 -renormalizations $(f_{\mathbf{a}} : U'_{\mathbf{a}} \rightarrow U_{\mathbf{a}})_{\mathbf{a} \in \Lambda}$ that form an analytic family of quadratic-like mappings. In particular, $\chi_1^{-1}(c_\infty)$ is an analytic set in \mathcal{Q}_1 .*

Of course a similar statement about \mathcal{Q}_2 is true, and it follows that \mathcal{Q}_1 and \mathcal{Q}_2 are open sets.

Proof. Let $\mathbf{a}_0 \in \mathcal{Q}_1$. Then there are topological disks U, U' with $\omega_1(\mathbf{a}_0) \in U'$ such that $f_{\mathbf{a}_0} : U' \rightarrow U$ is quadratic-like. Shrinking U and taking for U' the component of $f_{\mathbf{a}_0}(U)$ that is compactly contained in U , we may assume that the boundaries of U and U' are simple closed analytic curves and that the critical values $f_{\mathbf{a}_0}(\omega_i(\mathbf{a}_0))$, $i = 1, 2$, are not contained in ∂U . Choose a bi-disk Λ around \mathbf{a}_0 such that for all \mathbf{a} in a neighborhood of $\overline{\Lambda}$, the critical values avoid ∂U . We then have a unique holomorphic motion $h : \Lambda \times \partial U' \rightarrow \mathbb{C}$ of $\partial U'$ such that $f_{\mathbf{a}}(h(\mathbf{a}, z)) = f_{\mathbf{a}_0}(z) \in \partial U$. Since the image of a compact set under a holomorphic motion depends continuously on the parameter (in the Hausdorff metric on compact subsets of the sphere), we may suppose that $h_{\mathbf{a}}(\partial U')$ is contained in U for all $\mathbf{a} \in \Lambda$, by shrinking Λ if necessary. Note that $h(\mathbf{a}, \partial U)$ is a Jordan curve for each $\mathbf{a} \in \Lambda$, and denote by $U'_{\mathbf{a}}$ the domain bounded by it. Then $f_{\mathbf{a}} : U'_{\mathbf{a}} \rightarrow U$ is a quadratic-like mapping for each $\mathbf{a} \in \Lambda$. Set $\mathcal{U}' = \{(\mathbf{a}, z) \mid \mathbf{a} \in \Lambda, z \in U'_{\mathbf{a}}\}$ and $\mathcal{U} = \Lambda \times U$. We claim that $(f_{\mathbf{a}} : U'_{\mathbf{a}} \rightarrow U_{\mathbf{a}})_{\mathbf{a} \in \Lambda}$ is an analytic family of quadratic-like mappings. To see this we must verify (1)–(3) in the definition.

Let $\phi : U \rightarrow \mathbb{D}$ be a homeomorphism. Then $\Phi(\mathbf{a}, z) = (\mathbf{a}, \phi(z))$ is a homeomorphism over Λ mapping \mathcal{U} onto $\Lambda \times \mathbb{D}$. Let $\psi_{\mathbf{a}} : \mathbb{D} \rightarrow U'_{\mathbf{a}}$ be the conformal map that takes 0 to $\omega_1(\mathbf{a})$ and whose derivative $\psi'_{\mathbf{a}}(0)$ is real and positive. Then $\Psi(\mathbf{a}, z) = (\mathbf{a}, \psi_{\mathbf{a}}(z))$ is a homeomorphism over Λ , mapping $\Lambda \times \mathbb{D}$ onto \mathcal{U}' . This shows (1).

Note that the closure of \mathcal{U}' in \mathcal{U} is $\hat{\mathcal{U}}' = \{(\mathbf{a}, z) \mid \mathbf{a} \in \Lambda, z \in \overline{U'_{\mathbf{a}}}\}$. This follows since $\overline{U_{\mathbf{a}}}$ depends continuously on $\mathbf{a} \in \Lambda$ in the Hausdorff topology. Let $A \subset \Lambda$ denote a compact set. Then the preimage of this set under the projection of $\hat{\mathcal{U}}'$ onto Λ is $\{(\mathbf{a}, z) \mid \mathbf{a} \in A, z \in \overline{U'_{\mathbf{a}}}\}$. Since $\overline{U'_{\mathbf{a}}}$ depends continuously on \mathbf{a} , this set is compact. This shows (2).

Clearly $f : (\mathbf{a}, z) \mapsto f_{\mathbf{a}}(z)$ is complex analytic. Note that $f(\mathbf{a}, z) = f_{\mathbf{a}}(z)$ maps $\hat{\mathcal{U}}' \setminus \mathcal{U}'$ onto $\Lambda \times \partial U$. Hence the preimage K' of a compact set $K \subset \mathcal{U}$ avoids $\hat{\mathcal{U}}' \setminus \mathcal{U}'$. Using that $\pi_\Lambda(K') = \pi_\Lambda(K)$, it follows that K' does not intersect the boundary of \mathcal{U}' . So K' is a closed and bounded subset of \mathbb{C}^2 . We conclude that f is proper, which finishes the proof of (3).

That $\chi_1^{-1}(c_\infty)$ is analytic in \mathcal{Q}_1 now follows directly by applying Corollary 2 on page 313 in [DH2] and Proposition 6. □

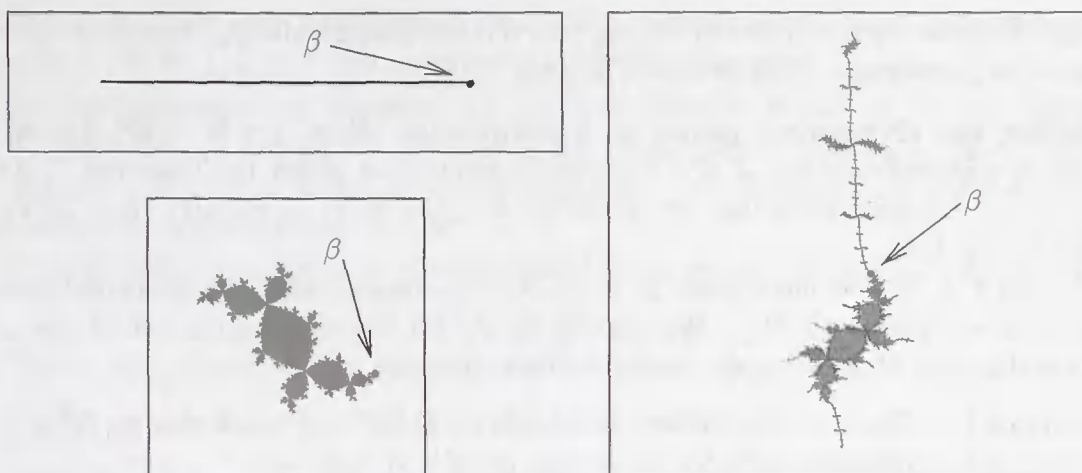


FIGURE 2. To the upper left is drawn an approximation of the Julia set of $z \mapsto z^2 - 2$. Below is the “rabbit”, that is, the Julia set of $z \mapsto z^2 + c$ with $c \approx -0.12236 + 0.74486i$. To the right is illustrated the Julia set of $f_{\mathbf{a}}(z) = z^3 - 3a^2z + b$, with $\mathbf{a} = (a, b) \approx (-0.03124478736 - 0.8615522432i, -0.1514859674 + 0.3334260962i)$. This cubic polynomial is the result of intertwining the two mentioned quadratic polynomials at their β -fixed points. Notice the copies of the Julia sets of the quadratic maps inside the Julia set of the cubic polynomial.

Next we are going to state a result from [EY] (see also [Ha]) that shows that Julia sets of quadratic polynomials show up inside Julia sets of certain cubic polynomials. To do so, we first need some notation. From the Straightening Theorem we know that for $\mathbf{a} \in \mathcal{Q}_i$, an ω_i -renormalization has a unique fixed point corresponding to the β -fixed point. This point we denote by $\beta_i(\mathbf{a})$.

The following result is an immediate consequence of the work of Epstein and Yampolsky in [EY]. They obtain it using intertwining surgery, although for this particular case the word “joining” surgery might be more appropriate since two quadratic-like mappings are joined together at their β -fixed points; see Figure 2.

Theorem 11. *There is a map $B : M \setminus \{1/4\} \times M \setminus \{1/4\} \rightarrow \mathcal{K}_1 \cap \mathcal{K}_2$, which is a homeomorphism onto its image, mapping (c_1, c_2) to \mathbf{a} , where \mathbf{a} satisfies $\beta_1(\mathbf{a}) = \beta_2(\mathbf{a})$ and $\chi_i(\mathbf{a}) = c_i$, $i = 1, 2$. \square*

We can now finish the proof of Theorem 7.

Let \mathcal{Q}'_1 denote the component of \mathcal{Q}_1 that contains $B(\{c_\infty\} \times (M \setminus \{1/4\}))$. The set $A = (\chi_1|_{\mathcal{Q}'_1})^{-1}(c_\infty)$ is analytic in \mathcal{Q}'_1 . This set cannot be of dimension 2. If it were, then it would contain a bi-disk, which contradicts that $B : (M \setminus \{1/4\}) \times (M \setminus \{1/4\}) \rightarrow \mathcal{Q}'_1$ is a homeomorphism. Since A contains the homeomorphic image of $M \setminus \{1/4\}$, it follows that A has dimension 1. In particular (see Remark 1), the set of singular points $S(A)$ is discrete in \mathcal{Q}'_1 .

Let $\mathcal{M} = \{c \in M \mid \exists k(P_c^k(0) = \beta(c))\}$, where $\beta(c)$ denotes the β -fixed point of $P_c : z \mapsto z^2 + c$. A standard normality argument (such as the one that shows that the boundary of M is contained in the closure of the center parameter, see e.g. [CG]) shows that this set is dense in ∂M . Now, no point in $\mathcal{M} \subset M$ is isolated, and we deduce that infinitely many points in the image $B(\{c_\infty\}, \mathcal{M})$ are regular.

Taking \mathbf{a}_0 for a regular point in $B(\{c_\infty\} \times \mathcal{M})$, the polynomial $f_{\mathbf{a}_0}$ has the required dynamical properties. This proves Theorem 7.

Relating the dynamical plane to a parameter slice. Let $\mathbb{D} \rightarrow \mathbb{C}^2, t \mapsto \mathbf{a}(t)$, denote a parametrization of the complex submanifold given by Theorem 7, with $\mathbf{a}(0) = \mathbf{a}_0$. Abusing notation, we write $f_t = f_{\mathbf{a}(t)}$, $\beta_1(t) = \beta_1(\mathbf{a}(t))$ and $\omega_i(t) = \omega_i(\mathbf{a}(t))$, $i = 1, 2$.

For all $t \in \mathbb{D}$, we have that f_t is ω_1 -renormalizable and the renormalization is hybrid equivalent to P_{c_∞} . We denote by $K^1(t)$ the filled Julia set of the ω_1 -renormalization of f_t . This set moves holomorphically with t :

Theorem 12. *There is a holomorphic motion $h : \mathbb{D} \times \mathbb{C} \rightarrow \mathbb{C}$ such that $h_t(K^1(0)) = K^1(t)$ and h_t commutes with the dynamics on $K^1(0)$, i.e.,*

$$h_t \circ f_0 = f_t \circ h_t$$

on $K^1(0)$.

Proof. Choose a period p and denote by $Z_p(t)$ the set of points in $K^1(t)$ that are periodic of period p under f_t .

Let $\mathcal{Y}_p = \{(t, z) \mid t \in \mathbb{D}, z \in Z_p(t)\}$, and let $\pi_i : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$ denote the projections such that for every $\mathbf{a} \in \mathbb{C} \times \mathbb{C}$ we have that $\mathbf{a} = (\pi_1(\mathbf{a}), \pi_2(\mathbf{a}))$. We claim that the projection $\pi_1 : \mathcal{Y}_p \rightarrow \mathbb{D}, (t, z) \mapsto t$, is a covering map. Let $t_0 \in \mathbb{D}$ denote an arbitrary point. Any ω_1 -renormalization $f_{t_0} : U' \rightarrow U$ is hybrid equivalent to P_{c_∞} ; so $\#Z_p(t)$ is constant. The set $Z_p(t_0)$ contains no indifferent cycles, for the same reason. It then follows from the Implicit Function Theorem that there are a neighborhood W of t_0 and holomorphic functions $z_i : W \rightarrow \mathbb{C}, i = 1, \dots, \#Z_p(t_0)$, such that $z_i(t) \in Z_p(t)$ for all i . Shrinking W , we can assume that $z_i - z_j$ is non-vanishing for $i \neq j$. That $\#Z_p(t)$ is constant implies that $\bigcup_i \{(t, z) \mid t \in W, z \in z_i(W)\}$ is an open neighborhood of the fiber $\pi_1^{-1}(t_0)$. The claim now follows because for each i the map $\pi_1 : \{(t, z) \mid t \in W, z \in z_i(W)\} \rightarrow W$ is a homeomorphism.

Using that the projection $\pi_1 : \mathcal{Y}_p \rightarrow \mathbb{D}$ is a covering map, we can lift $\pi_1 : \mathbb{D} \times Z_p(0) \rightarrow \mathbb{D}$ to $\pi_1 : \mathcal{X}_p \rightarrow \mathbb{D}$ and obtain a mapping $H : \mathbb{D} \times Z_p(0) \rightarrow \mathcal{Y}_p$.

Then define $h : \mathbb{D} \times Z_p(0) \rightarrow \mathbb{C}$ by $h = \pi_2 \circ H$; i.e., such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{C} & \xleftarrow{\pi_2} & \mathcal{Y}_p \\ \uparrow h & \nearrow H & \downarrow \pi_1 \\ \mathbb{D} \times Z_p(0) & \xrightarrow{\pi_1} & \mathbb{D} \end{array}$$

The map h is a holomorphic motion: By the Implicit Function Theorem, $t \mapsto h(t, z)$ is holomorphic for each $z \in Z_p(0)$. Suppose $h_{t_0}(x) = h_{t_0}(y)$. Now $t \mapsto (t, h(t, x))$ is a lift of the identity to the map $\pi_1 : \mathcal{Y}_p \rightarrow \mathbb{D}$, and the same is true for $t \mapsto (t, h(t, y))$. So by uniqueness of liftings, $h(t, x) = h(t, y)$ for all t . Since $z \mapsto h(0, z)$ is injective, it follows that $x = y$. Thus $z \mapsto h(t, z)$ is injective for all $t \in \mathbb{D}$.

The same line of arguments shows that h commutes with the dynamics. Both the map $t \mapsto (t, f_t \circ h(t, x))$ and $t \mapsto (t, h(t, f_0(x)))$ are lifts of the identity map to $\pi_1 : \mathcal{Y}_p \rightarrow \mathbb{D}$ that agree for $t = 0$, and by uniqueness they agree for all t . Thus the equation

$$h(t, f_0(x)) = f_t \circ h(t, x)$$

holds for all $t \in \mathbb{D}$ and all $x \in Z_p(0)$.

Noting that for each t the sets $Z_1(t), Z_2(t), \dots$ are disjoint, we get a holomorphic motion $h : \mathbb{D} \times \bigcup_p Z_p(0)$ that commutes with the dynamics.

By the Straightening Theorem, $\bigcup_p Z_p(0)$ is dense in $K^1(0)$, so by the λ -lemma (see [MSS]) we can extend h continuously to a holomorphic motion $h : \mathbb{D} \times K^1(0) \rightarrow \mathbb{C}$. Now $h(t, K^1(0)) = K^1(t)$ by continuity; and, also by continuity, the extension commutes with the dynamics.

Finally, using Ślodkowski's Theorem (see [Sl]), we can extend h to $h : \mathbb{D} \times \mathbb{C} \rightarrow \mathbb{C}$. \square

Define the connectedness locus \mathcal{S} of the family $\{f_t\}_{t \in \mathbb{D}}$ by

$$\mathcal{S} = \{t \in \mathbb{D} \mid K(f_t) \text{ is connected}\}.$$

Remark 2. Using a theorem of McMullen (see [Mc3]), it can be shown that \mathcal{S} contains a quasiconformal image of the connectedness locus $M_d \subset \mathbb{C}$ of the family $\{z \mapsto z^d + c\}_{c \in \mathbb{C}}$, with the image of the boundary of M_d contained in the boundary of \mathcal{S} , for some $d \geq 2$.

We will not use or prove this fact; instead we will show that the connectedness locus also contains a copy of J_∞ . For this we need a map relating the parameter disk $\{t \in \mathbb{D}\}$ to the dynamical plane $\{z \in \mathbb{C}\}$ of f_0 .

Lemma 2. *Let $h_t(z) = h(t, z)$, and denote by k the preperiod of $\omega_2(0)$ under f_0 . Then the mapping $H : \mathbb{D} \rightarrow \mathbb{C}$, $t \mapsto h_t^{-1} \circ f_t^k(\omega_2(t))$, is quasiregular on domains compactly contained in \mathbb{D} and non-constant. In fact, H is K -quasiregular on $\{t < \rho\}$, $\rho < 1$, with $K = (1 + \rho)/(1 - \rho)$.*

Proof. Let us first show that H is not constant. If it were, then for all $t \in \mathbb{D}$ we would have $f_t^k(\omega_2(t)) = h(\beta(0)) = \beta(t)$, where the last equality is due to the fact that h commutes with the dynamics. However, by Theorem 11 we can find a sequence of parameters t_n , converging to 0, such that f_{t_n} is ω_2 -renormalizable and the critical point of the renormalization is not prefixed.

Let us now show that H is quasiregular on domains compactly contained in \mathbb{D} . To see this, we will use the fact that the time t map of a holomorphic motion is quasiconformal. Indeed, it follows from Proposition 5 of [D] (compare with the λ -lemma in [MSS]) that for $|t| < \rho < 1$ the mapping h_t is K -quasiconformal with $K = (1 + \rho)/(1 - \rho)$.

Taking the $\frac{\partial}{\partial \bar{t}}$ distributional derivate of the equation

$$h_t \circ H(t) = f_t^k(\omega_2(t)),$$

we get

$$\frac{\partial}{\partial \bar{t}} h_t|_{H(t)} + \frac{\partial}{\partial z} h_t|_{H(t)} \frac{\partial H}{\partial \bar{t}}(t) + \frac{\partial}{\partial \bar{z}} h_t|_{H(t)} \frac{\partial \bar{H}}{\partial \bar{t}} = \frac{\partial}{\partial \bar{t}} f_t^k(\omega_2(t)).$$

Since $t \mapsto h(t, z)$ and $t \mapsto f_t^k(\omega_2(t))$ are holomorphic functions, it follows that

$$\frac{\partial}{\partial z} h_t|_{H(t)} \frac{\partial H}{\partial \bar{t}}(t) + \frac{\partial}{\partial \bar{z}} h_t|_{H(t)} \frac{\partial \bar{H}}{\partial \bar{t}} = 0.$$

Rearranging and taking absolute values, we obtain

$$\left| \frac{\partial H / \partial \bar{t}}{\partial H / \partial t}(t) \right| = \left| \frac{\partial h_t / \partial \bar{z}}{\partial h_t / \partial z}(H(t)) \right|.$$

So H is K quasiregular on $\{|t| < \rho\}$, for $\rho < 1$, with $K = (1 + \rho)/(1 - \rho)$. \square

Main theorem. We can now prove the main theorem:

Main Theorem. *There exist two cubic polynomials f, g which have no indifferent cycles and which are not topologically conjugate on their Julia sets, but which nevertheless have the same rational lamination. In particular, f and g are not combinatorially rigid.*

Proof. We denote by $\psi : U' \rightarrow V'$ a hybrid equivalence that maps a neighborhood U' of $K^1(0)$ to a neighborhood V' of $K(P_{c_\infty})$ and conjugates $f_0 : U' \rightarrow U$ to P_{c_∞} . Using that ψ is a conjugacy and Proposition 5, we get that $\beta_1(0)$ is accumulated by arbitrary small copies of $\hat{J}_\infty = \psi^{-1}(J_\infty)$, each of which is being mapped bijectively to \hat{J}_∞ by an iterate of f_0 . Since H is quasiregular and not constant, it maps 0 to $\beta_1(0)$ with a local degree $d > 0$. The copies of \hat{J}_∞ are full and do not contain $\beta_1(0)$, and from this it follows that we can find a copy \hat{J}'_∞ and an integer $l > 0$ such that

- f_0^l maps \hat{J}'_∞ bijectively onto \hat{J}_∞ ; and
- $H^{-1}(\hat{J}'_\infty)$ intersected with a neighborhood of 0 is the disjoint union of d continua C_1, \dots, C_d such that H maps each C_i bijectively onto \hat{J}'_∞ .

By bijectivity there is exactly one point $t_1 \in C_1$ such that $P_{c_\infty}^l \circ \psi(t_1) = 0 \in J_\infty$. Since h commutes with the dynamics on $K^1(0)$, we get

$$(1) \qquad f_{t_1}^{k+l}(\omega_2(t_1)) = \omega_1(t_1).$$

By injectivity of $H|_{C_1}$, we get that if $t_2 \in C_1$ and $t_2 \neq t_1$, then $f_{t_1}^{k+1}(\omega_2(t_1)) \neq \omega_1(t_1)$.

We claim that $f = f_{t_1}$ and $g = f_{t_2}$ for $t_2 \in C_1 \setminus \{t_1\}$ will satisfy the requirements of the theorem. We must prove:

- (a) f and g do not possess indifferent cycles;
- (b) f and g are not topologically conjugate on their Julia sets; and
- (c) f and g have the same rational lamination, $\lambda_{\mathbb{Q}}(f) = \lambda_{\mathbb{Q}}(g)$.

Part (a) is an immediate consequence of the following lemma:

Lemma 3. *Suppose $t \in \mathbb{D}$. If there exists an integer $k > 0$ such that $f_t^k(\omega_2(t)) \in K^1(t)$, then*

- (1) *the filled Julia set of f_t has no interior: $K(f_t) = J(f_t)$, and*
- (2) *all cycles of f_t are repelling.*

Part (b). Suppose there exists a homeomorphism $\eta : J(f) \rightarrow J(g)$ that conjugates f to g on $J(f)$. By Lemma 3, $K(f) = J(f)$ and $K(g) = J(g)$. The critical points $\omega_i(t_1)$ are distinguished in $K(f)$ in that they are the only points that have only one preimage. The critical points $\omega_i(t_2)$ are likewise distinguished in $K(g)$. It follows, since η is a conjugacy, that η maps the two critical points of f onto the two critical points of g . By equation (1), we must have either $g^{k+l}(\omega_2(t_2)) = \omega_1(t_2)$ or $g^{k+l}(\omega_1(t_2)) = \omega_2(t_2)$. However, we have already seen that $g^{k+l}(\omega_2(t_2)) \neq \omega_1(t_2)$, and we cannot have $g^{k+l}(\omega_1(t_2)) = \omega_2(t_2)$, since all the forward images of $\omega_1(t_2)$ are contained in $K^1(t_2)$ and $\omega_2(t_2)$ is not an element in $K^1(t_2)$. This proves (b).

Part (c). To complete the proof we shall show that the polynomials f_t have the same rational lamination for all $t \in C_1$. We will show that every rational ray lands, and that the landing point depends holomorphically on t for values of t in a neighborhood of C_1 . This will follow from the following stability result (Proposition 2 of expos   8 in the Orsay Notes, [DH1]).

Proposition 13. *Let $P_0 \in \mathcal{P}_d$ and $\theta \in \mathbb{Q}/\mathbb{Z}$. Assume that $R_{P_0}(\theta)$ lands at the (pre)periodic point $\gamma_0 \in J(P_0)$ and that the corresponding cycle is repelling. If $P_0^i(\gamma_0)$ is not a critical point for any $i \geq 0$, then there is a neighborhood Λ of P_0 in \mathcal{P}_d such that*

- (1) *the ray $R_P(\theta)$ lands at a (pre)periodic repelling point $\gamma(P)$ for all $P \in \Lambda$;*
- (2) *the map $\Lambda \times \{\eta \geq 0\} \rightarrow \mathbb{C}$ that maps (P, η) to the point of potential η on the ray $R_P(\theta)$ is continuous and holomorphic in P .* \square

Indeed, fix a rational angle θ and a parameter $t_0 \in C_1$. By Proposition 1, the dynamical ray of angle θ lands at a (pre)periodic point that is either parabolic or repelling. We have seen that f_t has no indifferent cycles. So to apply Proposition 13 we just have to see that the (forward) critical orbits do not contain the landing point. This follows easily, since the forward orbits of the critical point $\omega_i(t_0)$ are trapped in a forward invariant region on which f_{t_1} is conjugate to $P_{c_\infty} : J_\infty \rightarrow J_\infty$. By Theorem 4, J_∞ contains no (pre)periodic points. So the critical orbits do not contain any (pre)periodic point such as the landing point of $R_{t_0}(\theta)$.

We conclude that the ray $R_t(\theta)$ keeps landing and moves holomorphically in a neighborhood of each point in C_1 .

Suppose that the rays of angles θ_1 and θ_2 land at distinct points, for a parameter $t_0 \in \mathbb{D}$. Because the landing points depend holomorphically on t in a neighborhood of t_0 , they will keep on landing at distinct points for all parameters in a neighborhood of t_0 . Conversely, suppose two rational rays land at the same point for a parameter $t_0 \in C_1$. Since the rays are disjoint and the rays including landing points depend holomorphically on t in a neighborhood of t_0 , we get from Hurwitz's Theorem that the two rays keep landing at the same point for all parameters in a neighborhood of t_0 . Noting that C_1 is connected, we conclude that if two rays land at the same point for a parameter $t_0 \in C_1$, they will do so for all parameters $t \in C_1$. So $\lambda_{\mathbb{Q}}(f_t) = \lambda_{\mathbb{Q}}(f_0)$ for all $t \in C_1$. \square

We still have to prove Lemma 3. To do so we will use the fact that the multiplier at an indifferent fixed point is a quasiconformal invariant.

Lemma 4. *Suppose the two holomorphic germs $g_1, g_2 : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$ have indifferent fixed points at the origin, and a quasiconformal germ conjugates g_1 to g_2 . Then the multiplier $g_1'(0)$ is equal to the multiplier $g_2'(0)$.* \square

The lemma is well known; in fact, Naïshul' [Nai] proves that the multiplier of an indifferent fixed point is even a topological invariant.

We now prove Lemma 3.

Proof. It follows from the classification of Fatou components that (2) implies (1); so we need only prove (2). First note that $K^1(t)$ does not contain any non-repelling cycles. Indeed, the ω_1 -renormalization of f_t is hybrid equivalent to P_{c_∞} . So if $K^1(t)$ contains an indifferent cycle, then also P_{c_∞} possesses such a cycle by Lemma 4. But Theorem 4 asserts that all the cycles of P_{c_∞} are repelling.

We distinguish two cases, supposing f_t has a non-repelling cycle.

Case 1: The cycle is contained in $J(f_t)$. Then it is indifferent. Since both critical orbits are captured by $K^1(t)$, in the sense that $f_t^n(\omega_1(t)) \in K^1(t)$ for all $n \geq 0$ and $f_t^n(\omega_2(t)) \in K^1(t)$ for all $n \geq k$, the cycle must be contained in $K^1(t)$. We have seen that this cannot be the case.

Case 2: The cycle is contained in the interior of $K(f_t)$. Again, since the orbits of the critical points are captured by $K^1(t)$ and (by Theorem 4) this set has no interior, the cycle cannot be attracting. By the classification of Fatou components, it is necessarily a Siegel cycle. The boundary of the cycle of Siegel disks is contained in the closure of the postcritical set, and thus in $K^1(t)$. Since $K^1(t)$ is a full set, the Siegel disks themselves are also contained in $K^1(t)$. As before, this is not the case. \square

Question. I pass on a question I was asked by McMullen during a presentation of the result of this paper.

Question 1. Is there a natural combinatorial object associated to a polynomial $f \in \mathcal{C}_d$, $d > 2$, that uniquely determines the dynamics of f on $J(f)$?

ACKNOWLEDGMENT

I thank Bodil Branner and Carsten Petersen for reading and commenting on a version of this paper. Likewise I am grateful to Xavier Buff and Adrian Douady for instructive discussions.

REFERENCES

- [BH1] B. BRANNER and J. HUBBARD, *The iteration of cubic polynomials, Part I: The global topology of parameter space*, Acta Math., **160** (1988), no 3-4, 143–206. MR **90d**:30073
- [CG] L. CARLESON and T. GAMELIN, *Complex Dynamics*, Springer-Verlag, (1993). MR **94h**:30033
- [D] A. DOUADY, *Prolongement de mouvements holomorphes (d'après Ślodkowski et autres)*, (French) [Extension of holomorphic motions (after Ślodkowski and others)] Séminaire Bourbaki, Vol. 1993/94. Astérisque No. 227 (1995), Exp. No. 775, 3, 7–20. MR **95m**:58104
- [DH1] A. DOUADY and J. H. HUBBARD, *Etude dynamique des polynômes complexes I and II*, Publ. Math. d'Orsay (1984-85). MR **87f**:58072a
- [DH2] A. DOUADY and J. HUBBARD, *On the dynamics of polynomial-like mappings*, Ann. Sci. École Norm. Sup. Paris (4) **18** (1985), 287–343. MR **87f**:58083
- [EY] A. L. EPSTEIN and M. YAMPOLSKY, *Geography of the Cubic Connectedness Locus: Inter-twining Surgery*, Ann. Sci. École Norm. Sup. Paris (4), **32** (1999), no. 2, 151–185. MR **2000i**:37067
- [GF] H. GRAUERT and K. FRITZSCHE, *Several Complex Variables*, Springer-Verlag, (1976). MR **54**:3004
- [Ha] P. HAÏSSINSKY, Thèse de doctorat, Université Paris-Sud in Orsay, (1998).
- [KK] L. KAUP and B. KAUP, *Holomorphic Functions of Several Variables*, Walter de Gruyter, (1983). MR **85k**:32001
- [L1] M. LYUBICH, *The quadratic family as a qualitatively solvable model of chaos*, Notices Amer. Math. Soc. **47** (2000), no. 9, 1042–1052. MR **2001g**:37063
- [L2] M. LYUBICH, *Dynamics of quadratic polynomials. I, II*. Acta Math. **178** (1997), no. 2, 185–247, 247–297. MR **98e**:58145
- [MSS] R. MAÑÉ, P. SAD and D. P. SULLIVAN, *On the dynamics of rational maps*, Ann. Sci. École Norm. Sup. Paris (4) **16** (1983), no. 2, 193–217. MR **85j**:58089
- [M1] J. MILNOR, *Dynamics in one complex variable, Introductory Lectures*, Vieweg, 1999. MR **2002i**:37057
- [M2] J. MILNOR, *Local connectivity of Julia sets: Expository Lectures*, The Mandelbrot Set, Theme and Variations, edited by Tan Lei, Cambridge University Press (2000), pp. 67–116. MR **2001b**:37073
- [Mc1] C. T. MCMULLEN, *The Classification of Conformal Dynamical Systems*, Current Developments in Mathematics, 1995 (Cambridge, MA), 323–360, Internat. Press, Cambridge, MA, (1994). MR **98h**:58162
- [Mc2] C. T. MCMULLEN, *Complex Dynamics and Renormalization*, Annals of Mathematical Studies 135, Princeton University Press, (1994). MR **96b**:58097

- [Mc3] C. T. MCMULLEN, *The Mandelbrot Set is Universal*, The Mandelbrot Set: Theme and Variations, edited by Tan Lei, (2000), pp. 1–17. MR **2002f**:37071
- [Nai] V. A. NAISHUL', *Topological invariants of analytic and area preserving mappings and their applications to analytic differential equations in \mathbb{C}^2 and \mathbb{CP}^2* , Trans. Moscow Math. Soc. **42** (1983), 239–250. MR **84f**:58092
- [Sch] D. SCHLEICHER, *On fibers and local connectivity of Mandelbrot and Multibrot Sets*. Manuscript (1998).
- [Sl] Z. SŁODKOWSKI, *Extensions of holomorphic motions*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **22** (1995), 185–210. MR **96k**:30026
- [Sø] D. E. K. SØRENSEN, *Infinitely renormalizable quadratic polynomials, with non-locally connected Julia set*. J. Geom. Anal. **10** (2000), no 1, 169–206. MR **2001e**:37057

UNIVERSITÉ PAUL SABATIER, LABORATOIRE EMILE PICARD, 118, ROUTE DE NARBONNE, 31062
TOULOUSE CEDEX, FRANCE

E-mail address: `chris@picard.ups-tlse.fr`

Current address: Department of Mathematics, Technical University of Denmark, Matematik-
torvet, building 303, DK - 2800 Kgs Lyngby, Denmark

E-mail address: `christian.henriksen@mat.dtu.dk`

ZERO ENTROPY, NON-INTEGRABLE GEODESIC FLOWS AND A NON-COMMUTATIVE ROTATION VECTOR

LEO T. BUTLER

ABSTRACT. Let \mathfrak{g} be a 2-step nilpotent Lie algebra; we say \mathfrak{g} is *non-integrable* if, for a generic pair of points $p, p' \in \mathfrak{g}^*$, the isotropy algebras do not commute: $[\mathfrak{g}_p, \mathfrak{g}_{p'}] \neq 0$. *Theorem:* If G is a simply-connected 2-step nilpotent Lie group, $\mathfrak{g} = \text{Lie}(G)$ is non-integrable, $D < G$ is a cocompact subgroup, and \mathbf{g} is a left-invariant Riemannian metric, then the geodesic flow of \mathbf{g} on $T^*(D \backslash G)$ is neither Liouville nor non-commutatively integrable with C^0 first integrals. The proof uses a generalization of the rotation vector pioneered by Benardete and Mitchell.

1. INTRODUCTION

Let Σ be a C^∞ manifold, let $\mathbf{Z}(\Sigma)$ denote the set of C^∞ Riemannian metrics on Σ with zero topological entropy, and let $\mathbf{I}(\Sigma)$ denote those metrics with an integrable geodesic flow. Following Paternain [18], we can ask the questions:

Question A: *Is the set of integrable geodesic flows $\mathbf{I}(\Sigma)$ nonempty? Is the set of zero-entropy geodesic flows $\mathbf{Z}(\Sigma)$ nonempty?*

Question B: *What is the topological structure of $\mathbf{I}(\Sigma)$, resp. $\mathbf{Z}(\Sigma)$?*

The recent example in [3] by Bolsinov and Taĭmanov showed that $\mathbf{I}(\Sigma) \not\subseteq \mathbf{Z}(\Sigma)$. In this paper, it is shown that the reverse inclusion is false, too: $\mathbf{Z}(\Sigma) \not\subseteq \mathbf{I}(\Sigma)$. Thus, zero topological entropy is neither a necessary nor a sufficient condition for integrability.

There are well-known topological obstructions to the existence of a zero-entropy geodesic flow on a compact manifold: exponential word growth of the fundamental group and exponential growth of the Betti numbers of the loop space (over any field) [18]. An interesting class of manifolds where neither of these two well-known obstructions is effective is the class of *nilmanifolds*: quotients of nilpotent Lie groups. The fundamental groups of these manifolds are of polynomial word growth [1], and they are aspherical; so the loop space has trivial homology in all but the 0-th group. This paper constructs 2-step nilmanifolds on which all left-invariant geodesic flows are non-integrable, but their entropy is zero. It is unclear if these manifolds admit any geodesic flows that are integrable. It seems likely that they do not.

Received by the editors May 13, 2002 and, in revised form, September 23, 2002.

2000 *Mathematics Subject Classification.* Primary 37J30, 37E45; Secondary 53D25.

Key words and phrases. Rotation vector, geodesic flows, entropy, nilmanifolds, nonintegrability.

Research partially supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. Thanks to Gabriel Paternain, John Franks and Queen's University.

There has been some work related to the theme of this paper. Eberlein, Lee and Park, and Mast (amongst others) have studied the question of whether a left-invariant Riemannian metric \mathbf{g} on the two-step nilpotent Lie group G has a dense set of periodic points on $S(\Gamma \backslash G)$ for all lattice subgroups Γ [10], [11], [15], [17]. One way to understand this work is that the authors use the integrability of the geodesic flows in question to prove their results; the use of the formalism of the Euler-Lagrange equations on TG rather than Hamilton's equations on T^*G obscures this fact, however. In [4], [5], [6], explicit examples of integrable geodesic flows are constructed on a number of families of 2-step nilmanifolds including those studied in [10], [11], [15]. In [6] it is also shown that every left-invariant geodesic flow on every 2-step nilmanifold has zero topological entropy. The paper [7] constructs completely integrable, zero-entropy geodesic flows on an n -step filiform¹ nilmanifold for all n , thus proving that neither the step length nor the growth of the step length relative to the dimension of the group is an obstruction to the existence of integrable or zero-entropy geodesic flows. Finally, in [8], examples are constructed of n -step nilmanifolds, $n \geq 3$, with left-invariant metrics that have positive topological entropy.

The present paper constructs explicit examples of 2-step nilmanifolds whose left-invariant geodesic flows are not integrable.

Let us formulate the results of this paper more precisely. Let \mathfrak{g} be a Lie algebra and $p \in \mathfrak{g}^*$; denote by $\mathfrak{g}_p = \{x \in \mathfrak{g} : \text{ad}_x^* p = 0\}$. The point $p \in \mathfrak{g}^*$ is *regular* if $\dim \mathfrak{g}_p$ is minimal; a pair of points $p, p' \in \mathfrak{g}^*$ is *generic* if $\dim[\mathfrak{g}_p, \mathfrak{g}_{p'}]$ is minimal.

Definition 1.1. Let \mathfrak{g} be a 2-step nilpotent Lie algebra; \mathfrak{g} will be said to be *non-integrable* if, for a dense set of generic pairs of points $p, p' \in \mathfrak{g}^*$, we have $[\mathfrak{g}_p, \mathfrak{g}_{p'}] \neq 0$.

Lemma 3.2 proves that Definition 1.1 is equivalent to the existence of a generic pair of points $p, p' \in \mathfrak{g}^*$ such that $[\mathfrak{g}_p, \mathfrak{g}_{p'}] \neq 0$. A notion related to non-integrability is *almost non-singularity*: \mathfrak{g} is almost non-singular if there exists a regular point $p \in \mathfrak{g}^*$ such that the 2-form dp defined for all $x, y \in \mathfrak{g}$ by $dp(x, y) = -\langle p, [x, y] \rangle$ induces a symplectic form on $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ [15]. If \mathfrak{g} is almost non-singular, then for all regular points $p \in \mathfrak{g}^*$, $\mathfrak{g}_p = Z(\mathfrak{g})$. Non-integrable 2-step nilpotent Lie algebras cannot be almost non-singular 2-step nilpotent Lie algebras. There are, however, 2-step nilpotent Lie algebras that are neither almost non-singular nor non-integrable. The example of $\mathfrak{g} = \text{span} \{x, y_1, \dots, y_n, z_1, \dots, z_n\}$ with the relations $[x, y_i] = z_i$ is studied in [4], [5].

Here are two elementary constructions of a non-integrable 2-step nilpotent Lie algebra. Let \mathfrak{h} be a real, semi-simple Lie algebra and let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{h}$ be the vector space direct sum of 2 copies of \mathfrak{h} , with Lie bracket defined by $[x \oplus y, x' \oplus y'] = 0 \oplus [x, x']$. The example to keep in mind is $\mathfrak{h} = \mathfrak{so}(3)$. A second example is this: let V be a real, odd-dimensional vector space and let $\Lambda^2(V)$ be the vector space of alternating forms on V^* and let $\mathfrak{g} = V \oplus \Lambda^2(V)$ with Lie bracket $[x \oplus y, x' \oplus y'] = 0 \oplus x \wedge x'$. The lowest dimension for a non-integrable 2-step nilpotent Lie algebra is five: let $\mathfrak{g}_6 = \mathbf{R}^3 \oplus \Lambda^2(\mathbf{R}^3)$ and let $Z \subset Z(\mathfrak{g}_6)$ be a one-dimensional subalgebra; then $\mathfrak{g}_5 := \mathfrak{g}_6/Z$ is also non-integrable. One can verify that there are no further examples in dimensions 3 or 4.

¹The step length of a nilpotent Lie algebra of dimension $d \geq 2$ is bounded by $d-1$; it is filiform if it attains this bound.

Definition 1.2 (cf. [14]). Let $\phi_t : M \rightarrow M$ be a C^0 flow. It is integrable if there exists an open dense subset $L \subset M$ such that L is a C^0 torus fibre bundle, and the fibres of L are ϕ_t -invariant.

The hypothesis that L is a C^0 torus fibre bundle means that there is a C^0 manifold B and L is a locally trivial fibre bundle over B with fibres $\simeq \mathbf{T}^k$. The most common way to obtain integrable systems is via the Liouville-Arnol'd-Nehorošev theorem for Hamiltonian systems.

Theorem 1.3 (Main Theorem). *Let G be a simply-connected 2-step nilpotent Lie group, $\mathfrak{g} = \text{Lie}(G)$ be non-integrable, $D < G$ be a cocompact subgroup, and \mathfrak{g} be a left-invariant Riemannian metric. Let $\Sigma = D \backslash G$. Then the geodesic flow of \mathfrak{g} on $M = T^*\Sigma$ is not integrable in the sense of Definition 1.2. In particular, it is neither Liouville nor non-commutatively integrable with C^0 first integrals.*

Let us observe that not all nilpotent Lie groups have discrete cocompact subgroups; the existence of such a subgroup is known to be equivalent to the existence of a \mathbf{Q} -structure on \mathfrak{g} [16].

1.1. Outline. The proof of Proposition 1.3 is surprisingly straightforward and is related to a generalization of a rotation vector to non-commutative groups. Schwartzman [20] (see also [19]) introduced the notion of an asymptotic homology class in $H_1(M; \mathbf{R})$ for any curve $c : \mathbf{R} \rightarrow M$. Recall that $H_1(M; \mathbf{R}) = \pi/[\pi, \pi] \otimes \mathbf{R}$, where $\pi = \pi_1(M)$. This asymptotic homology class was generalized by Benardete and Mitchell [2] – who were building on the work of Chen [9] – to an asymptotic homotopy class of the curve c . Let us sketch the definition of this homotopy class. This principally requires an explanation of where it lives. Assume that π is a finitely-generated group. Let $\delta_0 := \pi$ and $\delta_{n+1} := [\pi, \delta_n]$. In general, π/δ_n has torsion, but the radical of δ_n , $\Delta_n = \sqrt{\delta_n}$, is a normal subgroup of π such that $\pi^n := \pi/\Delta_n$ is torsion-free. It is known that π^n is finitely-generated, n -step nilpotent and torsion-free. By a theorem of Mal'cev [16], there is a connected, simply-connected n -step nilpotent Lie group N_n such that π^n is a discrete, cocompact subgroup of N_n . The group N_n is unique up to isomorphism. Note that $\pi^1 \simeq H_1(M; \mathbf{Z})/\text{Tor } H_1(M; \mathbf{Z})$ and that $N_1 \simeq H_1(M; \mathbf{R})$. This construction, called the Mal'cev completion of π , is a generalization of the standard tensor product. The asymptotic homotopy class defined by Benardete and Mitchell lives in the connected, simply-connected 2-step nilpotent Lie group N_2 . Let us note that the construction of the groups N_n and Δ_n is “pointed” because $\pi = \pi(M; q)$ is pointed; so we should use the notation $\{N_{n,q}, \Delta_{n,q}\}$. A particularly nice feature of Benardete and Mitchell's paper [2] is that they embed the family of groups-with-lattice-subgroups $\{N_{n,q}, \Delta_{n,q}\}_{q \in M}$ in a connected, simply-connected n -step nilpotent Lie group N_n with lattice Δ_n and then show that there are isomorphisms $\phi_q : N_{n,q} \rightarrow N_n$ (which satisfy $\phi_q(\Delta_{n,q}) = \Delta_n$) such that $\phi_{q'} \circ \phi_q^{-1}$ is an inner automorphism of N_n for all $q, q' \in M$. Of course, there are no canonical choices for the ϕ_q .

The precise definition of the asymptotic homotopy class in N_2 requires some work. There is no ergodic theorem for these cocycles; so it is unknown if the asymptotic homotopy class exists for almost all initial conditions. Because the interest is really in understanding the asymptotic homotopy classes for a flow, we have elected to follow Fried [12] and define a related notion: projective homotopy classes. We show that it is possible to put enough algebraic structure on these

asymptotic homotopy classes in order to derive an effective necessary condition for integrability (see Lemma 2.7). Combined with a few calculations in section 3, this necessary condition gives a proof of Proposition 1.3.

2. FREE PROJECTIVE ASYMPTOTIC HOMOTOPY CLASSES

Let us continue with the notation in the outline. Let N be an n -step nilpotent Lie group and define $\mathbb{P}N$ to be the set of all one-parameter subgroups of N . Recall that if N is a connected, simply-connected nilpotent Lie group, then, by identifying $N \simeq \text{Lie}(N)$ via the exponential map, we can identify $\mathbb{P}N$ with $\mathbb{P}\mathfrak{n} = \mathfrak{n}/\mathbf{R}$, where $\mathfrak{n} := \text{Lie}(N)$. $\mathbb{P}\mathfrak{n}$ is the set of all subspaces of \mathfrak{n} with dimension at most 1. We can equip $\mathbb{P}\mathfrak{n}$ with the topology induced by the projection map $\pi(x) = \mathbf{R}x$, which makes it into a compact, connected non-Hausdorff topological space. We will also denote the coset $\mathbf{R}x$ by \bar{x} . The set $\{\bar{0}\} \subset \mathbb{P}\mathfrak{n}$ will be said to be *trivial*. Let us state the following:

Lemma 2.1. *Let L be a real Lie algebra, $\mathbb{P}L = L/\mathbf{R}$. Then*

- 1) *if $\phi : L \rightarrow L$ is a linear transformation, there is a continuous map $\bar{\phi} : \mathbb{P}L \rightarrow \mathbb{P}L$ such that $\pi \circ \phi = \bar{\phi} \circ \pi$;*
- 2) *for all $\bar{x} \in \mathbb{P}L$ there is a continuous map $\overline{\text{ad}}_{\bar{x}} : \mathbb{P}L \rightarrow \mathbb{P}L$ such that for all $x \in \pi^{-1}(\bar{x})$: $\pi \circ \text{ad}_x = \overline{\text{ad}}_{\bar{x}} \circ \pi$.*

The proof of Lemma 2.1 is straightforward. We note that (1) implies that the adjoint representation of N on \mathfrak{n} descends to an action on $\mathbb{P}\mathfrak{n}$, while (2) shows that, if $L = \mathfrak{n}$ is nilpotent, then we can introduce a grading on $\mathbb{P}\mathfrak{n}$ by saying that $\bar{x} \in \mathbb{P}\mathfrak{n}^k$ iff $\overline{\text{ad}}_{\bar{x}}^k = \bar{0}$ and $\overline{\text{ad}}_{\bar{x}}^{k-1} \neq \bar{0}$. This grading is inherited from the grading of \mathfrak{n} by its derived subalgebras. Part (2) also implies that the *projective bracket* $[\bar{x}, \bar{y}]$ is well-defined; it is also well-defined to say that \bar{x} and \bar{y} commute. This is equivalent to $[x, y] = 0$ for all $x \in \bar{x}$ and $y \in \bar{y}$, which is equivalent to $[\exp(tx), \exp(sy)] = 1$ for all s and t .

We are interested in the case where $N = N_n$ is the n -th part of the Mal'cev completion of $\pi = \pi_1(M; q)$. Let $\Pi : N \rightarrow \mathbb{P}\mathfrak{n}$ be defined by $\Pi = \pi \circ \log$. We want to define:

Definition 2.2. Let $c : [0, \infty) \rightarrow N$ be a continuous curve, $c(0) = 1$. Let $C_t := \{\Pi(c(s)) : s \geq t\}$ and define

$$[c] = \bigcap_{t>0} \overline{C_t} \subset \mathbb{P}\mathfrak{n}$$

to be the projective homotopy class of c . If $c'(t) := gc(t)$ for some $g \in N$, then $[c'] := \overline{\text{Ad}}_g[c]$.

Note that $\bar{0}$ lies in any nonempty closed subset of $\mathbb{P}\mathfrak{n}$. So $[c]$ always contains the trivial subset $\{\bar{0}\}$. For this reason, let us define the following set mapping $\hat{\pi}(V) := \pi(V) \cup \{\bar{0}\}$ for any $V \subseteq \mathfrak{n}$.

We can compute $[c]$ as follows: put some norm $|\cdot|$ on \mathfrak{n} , and let

$$d(t) = \exp(|\log c(t)|^{-1} \log c(t))$$

if $c(t) \neq id$ and 0 otherwise. Then $[c] = [d]$. This shows that $\bar{0} \neq \bar{x} \in [c]$ iff there is a sequence $x_k \in \mathfrak{n}$ of norm 1 such that $d(t_k) = x_k$ and $x_k \rightarrow x$. Thus, the projective homotopy class is a generalization of the homology directions (a projective rotation

vector) of Fried. The following two lemmata might be interpreted to say that Definition 2.2 is “correct”.

Lemma 2.3. *Let $x \in \mathfrak{n}$, $g \in N$, $c(t) := e^{tx}$, $c'(t) := gc(t)$ and $c''(t) := gc(t)g^{-1}$. Then $[c] = \hat{\pi}(x)$ and $[c'] = [c''] = \hat{\pi}(\text{Ad}_g x)$.*

Proof. The map $\Pi : N \rightarrow \mathbb{P}\mathfrak{n}$ satisfies $\Pi(e^{tx}) = \pi(\log(e^{tx})) = \pi(x)$ for $t \neq 0$. Thus $C_t = \{\pi(sx) : s \geq t\} = \pi(x)$ and so $\overline{C}_t = \hat{\pi}(x)$. The remaining claims are clear. \square

Given a continuous map $\gamma : \mathbf{T}^1 \rightarrow M$ we can extend γ to a map $\mathbf{R} \rightarrow M$. Let this map be denoted by $\hat{\gamma}$.

Let N be a connected, simply-connected nilpotent Lie group, $D < N$ a discrete, cocompact subgroup and let $\Sigma = D \backslash N$. For $\sigma = Dg$ there is a (non-canonical, non-unique) identification of $\pi_1(\Sigma; \sigma)$ with $\pi_1(\Sigma; De) = D$ by the map $\gamma(t) \rightarrow \gamma(t)g^{-1}$.

Lemma 2.4. *Assume that $D < N$ is a discrete, cocompact subgroup and let $\Sigma = D \backslash N$. Let $\gamma \in \pi_1(\Sigma; \sigma)$ be a nontrivial homotopy class, $\sigma = Dg$ and let $\Gamma : \mathbf{R} \rightarrow N$ be a lift of $\hat{\gamma}$ based at g . If γ is homotopic to the loop $t \rightarrow Dge^{tx}$, then $[\Gamma] = \hat{\pi}(\text{Ad}_g x)$.*

Proof. Assume that $\sigma = De$, $\Gamma(0) = e$, and let $d = \Gamma(1) \in D$. By hypothesis $d \neq e$. Then, there is a nonzero $x \in \mathfrak{n}$ such that $d = e^x$. Let $c(t) = e^{tx}$; so that $c(n) = d^n$ for all $n \in \mathbf{Z}$. The map $q(t) = \Gamma(t)c(t)^{-1}$ is 1-periodic and continuous; so $\text{im } q$ is a compact subset of N . Since $\Gamma(t) = q(t)c(t)$, we can write that $\log \Gamma(t) = tx + p(t)$. The set $\text{im } p$ is also compact; so $\log \Gamma(t)/|\log \Gamma(t)| = x/|x| + O(t^{-1})$ is well-defined for all t sufficiently large. Then $[\Gamma] = \hat{\pi}(x)$.

In the general case, we non-canonically identify $\pi_1(\Sigma; Dg)$ with $\pi_1(\Sigma; De)$ via the map $\gamma(t) \rightarrow \gamma(t)g^{-1} =: \tilde{\gamma}(t)$. From the arguments of the previous paragraph we can assume, without loss of generality, that $\gamma(t) = Dge^{tx}$. Then $\tilde{\gamma}(t) = De^{t\text{Ad}_g x}$. Apply Lemma 2.3 and the results of the previous paragraph to obtain the conclusion. \square

The π_1 de Rham theorem of Benardete and Mitchell allows one to introduce a second notion, which is more suitable for flows. Let $\phi_t : M \rightarrow M$ be a C^1 semi-flow, and let $q \in M$ and $n \geq 1$ be fixed. The Mal'cev completion of $\pi_1(M; q)$ induces a map $f_{q,n} : C^0(\mathbf{R}, 0), (M, q) \rightarrow N_n$. Let us denote a projective homotopy class in $\mathbb{P}\mathfrak{n}$ by $[\cdot]_n$. We define the projective homotopy class of $G(t) = \phi_t(q)$ to be $[f_{q,n} \circ G]_n$ and denote this by $[q]_{n,\phi}$. In the sequel, we will drop the subscript n, ϕ when this is understood. Let us also introduce a convenient notation: if $V \subset \mathfrak{n}$ (resp. $W \subset \mathbb{P}\mathfrak{n}$), then $\mathcal{O}(V) := \{\text{Ad}_g v : g \in N, v \in V\}$ (resp. $\overline{\mathcal{O}}(W) = \{\overline{\text{Ad}_g} w : g \in N, w \in W\}$). It is clear that $\overline{\mathcal{O}}(\pi(V)) = \pi(\mathcal{O}(V))$.

Definition 2.5. The free projective asymptotic homotopy class of q is the set

$$\mathcal{F}_{n,\phi}(q) := \overline{\mathcal{O}}([q]_{n,\phi}).$$

By Theorem 3.1 of [2] and Lemma 2.1, if $p = \phi_{t_0}(q)$, then $[q] = \overline{\text{Ad}_g} [p]$ for some $g \in N_n$. Thus, the free projective asymptotic homotopy class is an invariant of the orbit $\{\phi_t(q)\}_{t \in \mathbf{R}}$. In addition, although the construction of Theorem 3.1 in [2] depends on the smoothness of the manifold M , the free projective homotopy classes are an invariant of the C^0 conjugacy class of the semi-flow ϕ_t . In certain cases, it is possible to see this directly: when M is a vector bundle over a nilmanifold, the free projective asymptotic homotopy classes can be defined directly from Definition 2.2 without recourse to Theorem 3.1 of [2]. Note that Lemma 2.4 implies that Definition

2.5 is also “correct”: if $\phi_t(q) = qe^{tx}$, then the projective asymptotic homotopy class of $c_q(t) = qe^{tx}$ is $[c_q] = \hat{\pi}(\text{Ad}_q x)$, while the free projective asymptotic homotopy class is $\mathcal{F}_\phi(q) = \overline{\mathcal{O}}([c_q]) = \bigcup_{g \in N} [c_g]$.

Let us now explore the relationship of $\mathcal{F}_{n,\phi}(q)$ with integrability.

Definition 2.6. Let $U, V \subset \mathbb{P}n$. We say that U, V *weakly commute* if there exist nontrivial $\bar{x} \in U, \bar{y} \in V$ such that $[\bar{x}, \bar{y}] = \bar{0}$.

Lemma 2.7. Let M be a compact manifold and $\phi_t : M \rightarrow M$ be a C^0 semi-flow. Assume that there is a path-connected ϕ_t -invariant set $U \subset M$ such that $\pi_1(U)$ is abelian. Then for all $u, u' \in U$ and all $n \geq 1$, either $\mathcal{F}_{n,\phi}(u)$ and $\mathcal{F}_{n,\phi}(u')$ weakly commute or at least one of $\mathcal{F}_{n,\phi}(u), \mathcal{F}_{n,\phi}(u')$ is trivial.

Proof. Let $i : U \rightarrow M$ denote the inclusion map; i maps $\pi_1(U; u)$ into an abelian subgroup of $\pi_1(M; i(u))$ for all $u \in U$. By the naturality of the Mal'cev completion of $\pi_1(M)$, for each $n \geq 1$, $i_*\pi_1(U; u)$ gets mapped to a discrete torsion-free abelian subgroup \mathbf{A}_n of Δ_n , and so there is an abelian subgroup $A_n \subseteq N_n$ such that $A_n = \exp(\text{span log}(\mathbf{A}_n))$. The positive-invariance of U implies that $\mathcal{F}_{n,\phi}(u)$ and $\mathcal{F}_{n,\phi}(u')$ are subsets of A_n . Therefore, if there are nontrivial elements in both $\mathcal{F}_{n,\phi}(u)$ and $\mathcal{F}_{n,\phi}(u')$, then there are nontrivial elements in each set that commute; so the sets weakly commute. Otherwise, at least one of the sets is trivial. \square

Corollary 2.8. Let ϕ_t be a C^0 semi-flow. Assume that there is a C^0 embedding i of $U \simeq \mathbf{T}^k \times \mathbf{D}^l$ into M , such that $i(U)$ is ϕ_t -invariant. Assume that there is a C^0 map $\omega : \mathbf{D}^l \rightarrow \mathbf{R}^k$ such that for all $\theta \in \mathbf{T}^k, I \in \mathbf{D}^l$ we have $i^{-1}\phi_t i(\theta, I) = (\theta + t\omega(I), I)$. Then for all $u, u' \in U$ and all $n \geq 1$, either the sets $\mathcal{F}_{n,\phi}(i(u))$ and $\mathcal{F}_{n,\phi}(i(u'))$ weakly commute or at least one is trivial.

3. TWO-STEP NILPOTENT LIE GROUPS

Let \mathfrak{g} be a 2-step nilpotent Lie algebra with center $\mathfrak{z} = Z(\mathfrak{g})$, so that $[\mathfrak{g}, \mathfrak{g}] \subset Z(\mathfrak{g})$, let \langle, \rangle be an inner product on \mathfrak{g} , and let

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{z}$$

be an \langle, \rangle -orthogonal decomposition of \mathfrak{g} . The Lie bracket on \mathfrak{g} is written as $[x + y, x' + y'] = [x, x']$ for all $x, x' \in \mathfrak{h}$ and $y, y' \in \mathfrak{z}$, and so the commutator defines a skew-symmetric, bilinear form $\omega : \mathfrak{h} \times \mathfrak{h} \rightarrow \mathfrak{z}$ by $\omega(x, x') = [x, x']$.

The Lie algebra \mathfrak{g} can also be given the structure of a Lie group $(G, *)$ by $X * Y := X + Y + \frac{1}{2}[X, Y]$, so that $\mathfrak{g} = \text{Lie}(G)$ and the exponential map is the identity. In the sequel, elements in G will often be viewed as elements in \mathfrak{g} under the inverse (logarithm) map – which is the identity map in these coordinates. If D is a discrete, cocompact subgroup of G , then there exists a generating set $X_1, \dots, X_p, Y_1, \dots, Y_q$ where Y_1, \dots, Y_q generate $Z(D)$ and the cosets $X_1 + Z(D), \dots, X_p + Z(D)$ generate $D/Z(D)$ and $p = \dim \mathfrak{h}, q = \dim \mathfrak{z}$ [16]. The generating set therefore determines a basis of \mathfrak{g} and an inner product \langle, \rangle' relative to which it is an orthonormal basis. It may be supposed then that $\langle, \rangle = \langle, \rangle', \mathfrak{h} = \text{span}_{\mathbf{R}}\{X_1, \dots, X_p\}$ and $\mathfrak{z} = \text{span}_{\mathbf{R}}\{Y_1, \dots, Y_q\}$.

Lemma 3.1. Let $D \leq G$ be a discrete, cocompact subgroup and let $(,)$ be an inner product on \mathfrak{g} . Then there exists an automorphism $f : G \rightarrow G$ and a subgroup $D' = f^{-1}(D)$ with generators $X_1, \dots, X_p, Y_1, \dots, Y_q$ such that $(X_i, Y_j) = 0$. In

addition, if \mathfrak{g} is the left-invariant metric on G determined by (\cdot, \cdot) , then $(D' \setminus G, f^* \mathfrak{g})$ is isometric to $(D \setminus G, \mathfrak{g})$.

Proof. Let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{z}$ be the $\langle \cdot, \cdot \rangle$ -orthogonal decomposition of \mathfrak{g} . Let $a(x)$ be the $\langle \cdot, \cdot \rangle$ -orthogonal projection of $x \in \mathfrak{h}$ onto \mathfrak{z} . The map $F : x + y \rightarrow x - a(x) + y$ for all $x \in \mathfrak{h}$ and $y \in \mathfrak{z}$ is an automorphism of \mathfrak{g} . Let $f = \exp \circ F \circ \log$ be the map induced by F on G . Then f is an automorphism and by construction $F(\mathfrak{h})$ is $\langle \cdot, \cdot \rangle$ -orthogonal to \mathfrak{z} . \square

Lemma 3.1 is proven in [13] for Heisenberg groups. The importance of this lemma is that, by fixing a discrete, cocompact subgroup D with a fixed generating set, attention can be confined to those metrics that are block diagonal relative to this fixed basis of \mathfrak{g} . Here and henceforth, $\langle \cdot, \cdot \rangle$ will be a fixed inner product on \mathfrak{g} relative to which $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{z}$, D will be a discrete, cocompact subgroup of G with $\langle \cdot, \cdot \rangle$ -orthonormal generating set $X_1, \dots, X_p, Y_1, \dots, Y_q$ and (\cdot, \cdot) will be a second inner product that is block diagonal: for all $X, X' \in \mathfrak{h}$ and $Y, Y' \in \mathfrak{z}$,

$$(3.1) \quad (X + Y, X' + Y') = \langle X, \mathcal{A}X \rangle + \langle Y, \mathcal{B}Y' \rangle$$

where $\mathcal{A}_{ij} = (X_i, X_j)$ and $\mathcal{B}_{kl} = (Y_k, Y_l)$. The metric \mathfrak{g} on G will be the left-invariant metric determined by (\cdot, \cdot) or, equivalently, the pair \mathcal{A}, \mathcal{B} .

Finally, let \mathfrak{g} be a real, nonabelian Lie algebra and let $r = \min_{p \in \mathfrak{g}^*} \dim \mathfrak{g}_p$; r is commonly called the *index* of \mathfrak{g} . It is well-known that the set of p where $\dim \mathfrak{g}_p > r$ is a nontrivial algebraic set; let \mathfrak{g}_r^* denote the complement of this set. Let G_r denote the Grassmannian of r -dimensional subspaces of \mathfrak{g} , and let $\mathbf{G}_r = G_r \times \mathfrak{g}_r^*$ denote the trivial G_r bundle over \mathfrak{g}_r^* . The map $p \rightarrow \mathfrak{g}_p$ restricted to \mathfrak{g}_r^* is a section of \mathbf{G}_r . Since $x \in \mathfrak{g}_p$ iff $\text{ad}_x^* p = 0$ iff $\sum_{\alpha, \beta=1}^n c_{\beta, i}^\alpha p_\alpha x^\beta = 0$ for $i = 1, \dots, n$ (where $c_{\beta, i}^\alpha$ are the structure constants of \mathfrak{g}) iff x satisfies a system of linear equations with coefficients linear functions of p , so it follows that the section $p \rightarrow \mathfrak{g}_p$ is algebraic in p . Thus, the map $f : \mathfrak{g}_r^* \rightarrow \mathfrak{g}$ defined by $p \rightarrow [\mathfrak{g}_p, \mathfrak{g}_p]$ is algebraic and so the subset $\{f \neq 0\}$ is a Zariski-open subset of \mathfrak{g}_r^* . Consequently, $\{f \neq 0\}$ is either empty or else it is dense in \mathfrak{g}_r^* . We have therefore proven:

Lemma 3.2. *Let \mathfrak{g} be a 2-step nilpotent Lie algebra. Then \mathfrak{g} is nonintegrable iff there exists two generic points $p, p' \in \mathfrak{g}_r^*$ such that $\dim [\mathfrak{g}_p, \mathfrak{g}_{p'}] > 0$.*

3.1. Geodesic equations of motion. Let \mathfrak{g} be a 2-step nilpotent Lie algebra. Let $A : \mathfrak{z}^* \rightarrow \text{so}(\mathfrak{h})$ be defined for all $x, x' \in \mathfrak{h}$ and $q \in \mathfrak{z}^*$ by $\langle x, A(q)x' \rangle := \langle q, [x, x'] \rangle$. Observe that

Lemma 3.3. $\ker A(q) = \mathfrak{h} \cap \mathfrak{g}_q$.

Proof. $x \in \ker A(q) \subset \mathfrak{h}$ iff $0 = \langle q, [x, x'] \rangle = -\langle \text{ad}_x^* q, x' \rangle$ for all $x' \in \mathfrak{h}$. Since $\mathfrak{g} = \mathfrak{h} \oplus Z(\mathfrak{g})$, this is true iff $0 = \langle q, [x, x'] \rangle = -\langle \text{ad}_x^* q, x' \rangle$ for all $x' \in \mathfrak{g}$. \square

Lemma 3.4. *\mathfrak{g} is nonintegrable iff there exists generic points $q, q' \in \mathfrak{z}^*$ such that $[\ker A(q), \ker A(q')] \neq 0$.*

Proof. Apply Lemmas 3.2 and 3.3. \square

Let (x, y, p, q) be the coordinates of a point in $T^*G = \mathfrak{h} \times \mathfrak{z} \times \mathfrak{h}^* \times \mathfrak{z}^*$. The Hamiltonian of the metric \mathfrak{g} on T^*G is $H_{\mathfrak{g}} = \frac{1}{2} \langle p, Rp \rangle + \frac{1}{2} \langle q, Sq \rangle$ where $R = \mathcal{A}^{-1}$ and $S = \mathcal{B}^{-1}$. The equations of motion are

$$(3.2) \quad X_{H_{\mathfrak{g}}} = \begin{cases} \dot{q} &= 0, & \dot{y} &= Sq + \frac{1}{2}[x, Rp], \\ \dot{p} &= -A(q)Rp, & \dot{x} &= Rp. \end{cases}$$

Then q is a \mathfrak{z}^* -valued first integral of X_{H_g} and $F := p + A(q)x$ is an \mathfrak{h}^* -valued first integral. One way to think about the integrability of X_{H_g} is to determine in which cases F can be "pushed down" to $T^*(D \setminus G)$.

Let $r^2 = R$ be a positive-definite, symmetric square root of R (recall that \mathfrak{h}^* is identified with \mathfrak{h} via the inner product \langle, \rangle). Let $B(q) = rA(q)r \in so(\mathfrak{h})$ and $p = rv$, which implies $\dot{v} = -B(q)v$. Because $B(q) \in so(\mathfrak{h})$, both $\ker B(q)$ and its \langle, \rangle -orthogonal complement are $B(q)$ invariant. Let $K_q = \ker B(q)$ and $L_q = K_q^\perp$. Then for each $q \in \mathfrak{z}^*$, $\mathfrak{h} = K_q \oplus L_q$. Note that $K_q = r^{-1} \ker A(q)$.

Let $k = \inf_{q \in \mathfrak{z}^*} \dim K_q$ and $l = \sup_{q \in \mathfrak{z}^*} \dim L_q$, let $G_s(\mathfrak{h}^*)$ denote the Grassmannian of s -dimensional planes in \mathfrak{h}^* and let $G_s = G_s(\mathfrak{h}^*) \times \mathfrak{z}^*$ denote the trivial bundle. The following is obvious: there is a subset $\mathfrak{z}_r^* \subset \mathfrak{z}^*$ such that $\mathfrak{z}^* - \mathfrak{z}_r^*$ is a closed algebraic set and the map $q \rightarrow K_q$ (resp. $q \rightarrow L_q$) is an algebraic section of $G_k|_{\mathfrak{z}_r^*}$ (resp. $G_l|_{\mathfrak{z}_r^*}$).

For each $v \in \mathfrak{h}^*$ write $v = v_0 + v_1$ where $v_0 = v_0(q) \in K_q$ and $v_1 = v_1(q) \in L_q$. Abuse notation, and let $B(q)^{-1} = (B(q)|_{L_q})^{-1}$ when $K_q \neq \mathfrak{h}^*$. We can integrate the equations 3.2 to obtain

$$(3.3) \quad \phi_t(x, y, v, q) = \begin{cases} q(t) &= q, & y(t) &= y + tSq + \frac{1}{2} \int_0^t [x(s), rv(s)] ds, \\ v(t) &= e^{-tB(q)}v, & x(t) &= x + trv_0 + rB(q)^{-1}\{1 - e^{-tB(q)}\}v_1. \end{cases}$$

Note that $x(t) = trv_0 + O(1) = tu + O(1)$, where $u = rv_0 \in \ker A(q)$. The expression for $y(t)$ may be expanded to yield:

$$(3.4) \quad \begin{aligned} y(t) &= y + tSq + \frac{1}{2} \int_0^t \left\{ [x, rv_0] + [x, re^{-B(q)s}v_1] \right\} ds \\ &\quad + \int_0^t \left\{ s[rv_0, re^{-sB(q)}v_1] + [rB(q)^{-1}(1 - e^{-sB(q)})v_1, rv_0 + re^{-sB(q)}v_1] \right\} ds. \end{aligned}$$

Inspection of the integrands shows that the first, second and fourth integrands are bounded functions of s . So their integrals have a norm bounded by $\text{const.} \times t$. The substitution $u = srv_0$ and $dv = re^{-B(q)s}v_1 ds$ combined with integration by parts on the third integrand shows that its norm is bounded by $\text{const.} \times t$, also.

Lemma 3.5. *Let $P = (x, y, p, q) \in T^*G$ be such that $B(q) \neq 0$. Let $p = rv$ and write $v = v_0 + v_1$ (resp. $p = p_0 + p_1$) with $p_i = rv_i$. Assume that $v_0 \neq 0$. Then the projective asymptotic homotopy class of $\phi_t(P)$ is nontrivial and $[P]_{2,\phi} \subseteq \hat{\pi}(p_0 + \mathfrak{z})$.*

Proof. Let $\mathbf{p} : T^*G \rightarrow G$ denote the canonical projection. Assume that $(x, y) = (0, 0)$. Since $|y(t)| \leq \text{const.} \times t$, $v_0 \neq 0$, $x(t) = trv_0 + O(1)$ and $p_0 = rv_0$, we have that $\mathbf{p}\phi_t(P)/|\mathbf{p}\phi_t(P)| = c(t)p_0 + z(t)$, where $c(t)$ is a positive function of t bounded away from 0 and ∞ and $z(t) \in \mathfrak{z}$ has norm < 1 . Thus $\pi(\mathbf{p}\phi_t(P)) = \pi(p_0 + c(t)^{-1}z(t))$. Since $|c(t)^{-1}z(t)| \leq \text{const.}$, this shows that $[P]_{2,\phi} \subseteq \hat{\pi}(p_0 + \mathfrak{z})$.

Assume now that $g = (x, y) \neq (0, 0)$. The left-invariance of ϕ_t and the results of the previous paragraph show that $[P]_{2,\phi} \subseteq \hat{\pi}(\text{Ad}_g p_0 + \mathfrak{z})$. Since \mathfrak{g} is 2-step nilpotent, one computes that $\text{Ad}_g p_0 + \mathfrak{z} = p_0 + \mathfrak{z}$, which proves the lemma. \square

Corollary 3.6. *Assume the hypotheses of the previous lemma. Then the free projective asymptotic homotopy class of $\phi_t(P)$ is nontrivial and $\mathcal{F}_{2,\phi}(P) \subseteq \hat{\pi}(p_0 + \mathfrak{z})$.*

Proof. Apply the previous lemma and the definition of \mathcal{F} . \square

Proposition 3.7. *Assume that \mathfrak{g} is a nonintegrable 2-step nilpotent Lie algebra with a \mathbf{Q} -structure. Let $D < G$ be a discrete cocompact subgroup, $\Sigma = D \backslash G$ and \mathbf{g} be a left-invariant metric on G . Then the geodesic flow of \mathbf{g} on $T^*\Sigma$ is not integrable.*

Proof. Let $\hat{\phi}_t$ (resp. ϕ_t) denote the geodesic flow of \mathbf{g} on $T^*\Sigma$ (resp. T^*G). The geodesic flows satisfy $\hat{\phi}_t(DP) = D\phi_t(P)$ for all $P \in T^*G$. Let $DP, DP' \in T^*\Sigma$. From the definition of \mathcal{F} , $\mathcal{F}_{n,\hat{\phi}}(DP) = \mathcal{F}_{n,\phi}(P)$ and similarly for DP' . By Corollary 3.6, $\{0\} \neq \mathcal{F}_{2,\phi}(P) \subseteq \hat{\pi}(p_0(q) + \mathfrak{z})$ and $\{0\} \neq \mathcal{F}_{2,\phi}(P') \subseteq \hat{\pi}(p'_0(q') + \mathfrak{z})$. Let us remark that if $\bar{x} \in \mathcal{F}_{2,\phi}(P)$ and $\bar{x}' \in \mathcal{F}_{2,\phi}(P')$ are both nontrivial elements, then $[\bar{x}, \bar{x}'] = \pi([p_0(q), p'_0(q')])$. By Lemma 3.4 and the subsequent discussion, $p_0(q) \in \mathfrak{g}_q \cap \mathfrak{h}$ and $p'_0(q') \in \mathfrak{g}_{q'} \cap \mathfrak{h}$. By the hypothesis that \mathbf{g} is nonintegrable and 2-step nilpotent, it follows that for an open dense set of $(q, q') \in \mathfrak{z}^* \times \mathfrak{z}^*$ there exists $u_0 \in \mathfrak{g}_q \cap \mathfrak{h}$, $u'_0 \in \mathfrak{g}_{q'} \cap \mathfrak{h}$ such that $[u_0, u'_0] \neq 0$. Thus, for an open and dense set of $(DP, DP') \in T^*\Sigma \times T^*\Sigma$ the free projective asymptotic homotopy classes $\mathcal{F}_{2,\hat{\phi}}(DP)$ and $\mathcal{F}_{2,\hat{\phi}}(DP')$ are nontrivial and do not weakly commute.

Assume $\hat{\phi}_t : T^*\Sigma \rightarrow T^*\Sigma$ is integrable in the sense of Definition 1.2. Then there exists an open dense set $L \subset T^*\Sigma$ such that $L = \bigcup L_\alpha$ and the open sets L_α are $\hat{\phi}_t$ -invariant and have an abelian fundamental group. Thus, for an open dense set of points $DP \in T^*\Sigma$ there is an open neighbourhood of DP , $L_\alpha(DP) \subset L$, which is $\hat{\phi}_t$ -invariant and has an abelian fundamental group. Since the free projective asymptotic homotopy classes of an open dense set of points $(DP, DP') \in L_\alpha \times L_\alpha$ are nontrivial, Lemma 2.7 implies that $\mathcal{F}_{2,\hat{\phi}}(DP)$ and $\mathcal{F}_{2,\hat{\phi}}(DP')$ weakly commute. Therefore, there is an open set of points $(DP, DP') \in T^*\Sigma \times T^*\Sigma$ with weakly commuting free projective asymptotic homotopy classes. This contradicts the conclusion of the previous paragraph, which is absurd. \square

REFERENCES

1. H. Bass, *The degree of polynomial growth of finitely generated nilpotent groups*, Proc. London Math. Soc. (3) 25, 603–614 (1972). MR 52:577
2. D. Benardete and J. Mitchell, *Asymptotic homotopy cycles for flows and Π_1 de Rham theory*, Trans. Amer. Math. Soc. 338(2), 495–535 (1993). MR 93j:58107
3. A. V. Bolsinov and I. A. Taimanov, *Integrable geodesic flows with positive topological entropy*, Invent. Math. 140(3), 639–650 (2000). MR 2001b:37081
4. L. T. Butler, *A new class of homogeneous manifolds with Liouville-integrable geodesic flows*, C. R. Math. Acad. Sci. Soc. R. Can. 21(4), 127–131 (1999). MR 2001i:53141
5. L. T. Butler, *New examples of integrable geodesic flows*, Asian J. Math. 4(3), 515–526 (2000). MR 2001i:37090
6. L. T. Butler, *Integrable geodesic flows with wild first integrals: The case of two-step nilmanifolds*, Ergodic Theory Dynam. Systems, to appear.
7. L. T. Butler, *Integrable geodesic flows on n -step nilmanifolds*, J. Geom. Phys. 36(3–4), 315–323 (2000). MR 2002j:37077
8. L. T. Butler, *Invariant metrics on a nilmanifold with positive topological entropy*, submitted to Geometriae Dedicata. 2001.
9. K. T. Chen, *Extension of C^∞ function algebra by integrals and Malcev completion of π_1* , Advances in Math. 23(2), 181–210 (1977). MR 56:16664
10. P. Eberlein, *Geometry of 2-step nilpotent groups with a left invariant metric*, Ann. Sci. École Norm. Sup. 27(5), 611–660 (1994). MR 95m:53059
11. P. Eberlein, *Geometry of 2-step nilpotent groups with a left invariant metric. II*, Trans. Amer. Math. Soc. 343(2), 805–828 (1994). MR 95b:53061
12. F. Fried, *The geometry of cross sections to flows*, Topology 21(4), 353–371 (1982). MR 84d:58068

13. C. S. Gordon and E. N. Wilson, *The spectrum of the Laplacian on Riemannian Heisenberg manifolds*, Michigan Math. J. 33(2), 253–271 (1986). MR **87k**:58275
14. A. B. Katok, *Ergodic perturbations of degenerate integrable Hamiltonian systems*, Izv. Akad. Nauk SSSR Ser. Mat. 37, 539–576 (1973). MR **48**:9758
15. K. B. Lee and K. Park, *Smoothly closed geodesics in 2-step nilmanifolds*, Indiana Univ. Math. J. 45(1), 1–14 (1996). MR **97h**:53044
16. A. I. Malcev, *On a class of homogeneous spaces*, Amer. Math. Soc. Translation, no. 39, 1951. MR **12**:589e
17. M. Mast, *Closed geodesics in 2-step nilmanifolds*, Indiana Univ. Math. J. 43, 885–911 (1994). MR **96a**:53057
18. G. P. Paternain, *Geodesic flows*, Progress in Math., vol. 180, Birkhäuser, Boston, MA, 1999. MR **2000h**:53108
19. F. Rhodes, *Asymptotic cycles for continuous curves on geodesic spaces*, J. London Math. Soc. (2), 6, 247–255 (1973). MR **49**:6208
20. S. Schwartzman, *Asymptotic cycles*, Ann. of Math. (2) 66, 270–284 (1957). MR **19**:568i

DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, 2033 SHERIDAN ROAD,
EVANSTON, ILLINOIS 60208

E-mail address: lbutler@math.northwestern.edu

COMPLETE HOMOGENEOUS VARIETIES: STRUCTURE AND CLASSIFICATION

CARLOS SANCHO DE SALAS

ABSTRACT. Homogeneous varieties are those whose group of automorphisms acts transitively on them. In this paper we prove that any complete homogeneous variety splits in a unique way as a product of an abelian variety and a parabolic variety. This is obtained by proving a rigidity theorem for the parabolic subgroups of a linear group. Finally, using the results of Wenzel on the classification of parabolic subgroups of a linear group and the results of Demazure on the automorphisms of a flag variety, we obtain the classification of the parabolic varieties (in characteristic different from 2, 3). This, together with the moduli of abelian varieties, concludes the classification of the complete homogeneous varieties.

0. INTRODUCTION

Let X be a variety over an algebraically closed field k . Let G be the functor of automorphisms of X that acts naturally on X . We say that X is *homogeneous* if G acts transitively on X , that is, if for each pair of points $x, x' \in \text{Hom}(S, X)$, there exists an automorphism $\tau \in \text{Aut}_S(X \times S)$ (after a faithfully flat base change on S) transforming one into the other: $\tau(x) = x'$.

It is known that the group of automorphisms of a complete variety exists, i.e., the functor G is representable, and it is locally of finite type (see [7]). Hence, if X is smooth, connected and homogeneous, then the reduced connected component through the origin of G acts transitively on X . Therefore, every smooth and connected homogeneous variety is isomorphic to G/P , with G a smooth and connected algebraic group and $P \subset G$ a subgroup.

Let $\text{Aut}^0(X)$ denote the reduced connected component through the origin of the automorphism scheme of X .

The main results that we obtain here are:

(1) *A complete homogeneous variety splits canonically and uniquely as a direct product of an abelian variety and a parabolic variety.*

A parabolic variety means a variety of the form G/P with G an affine, smooth and connected algebraic group and P a parabolic subgroup (eventually not reduced), i.e., a subgroup containing a Borel subgroup of G . This is Theorem 5.2. The analogue of this result for compact Kähler manifolds is due to A. Borel and R. Remmer (see [2]).

Received by the editors February 15, 2002 and, in revised form, October 11, 2002.

2000 *Mathematics Subject Classification.* Primary 14M17, 14M15, 14L30, 32M10.

This research was partially supported by the Spanish DGI through research project BFM2000-1315 and by the "Junta de Castilla y León" through research project SA009/01.

For the proof, it is necessary to prove that any action of an affine group on a variety with projective orbits is trivial. More precisely, it is proved that *if an affine algebraic group acts on a variety and the orbits are projective and equidimensional, then the variety splits as the direct product of an orbit and the quotient by the action*. This is Theorem 3.3. Borel proved an analogue of this result for a complex variety (see [1]).

(2) From this result the next one, of great interest for the classification, follows easily: *The automorphism group of a homogeneous complete variety classifies (modulo isomorphisms) the variety, up to the choice, in such a group, of a class (modulo automorphisms of the group) of a parabolic subgroup $[P]$. Moreover, this group of automorphisms splits, in a unique way, as the direct product of a semisimple group of adjoint type (that is, with trivial center) and an abelian variety, and the subgroup P is a parabolic subgroup of the semisimple part*. That is, to each homogeneous variety X is assigned its automorphism group $\text{Aut}^0(X) = G \times A$ and the class $[P]$, modulo automorphisms, of a parabolic subgroup $P \subset G$; thus the maps $X \mapsto (G, A, [P])$ and $(G, A, [P]) \mapsto X = G/P \times A$ establish an equivalence of objects modulo isomorphisms. This is Theorem 5.7.

It should be noted that the triplets $(G, A, [P])$ are chosen with the single condition that $\text{Aut}^0(G/P) = G$. Therefore, a first question is to know whether G may be any semisimple group of adjoint type; in other words, if G is given, does there exist a parabolic subgroup $P \subset G$ such that $\text{Aut}^0(G/P) = G$? The answer is affirmative, and is due to Demazure ([4]). From his results one obtains in particular:

If G is a semisimple group of adjoint type and \mathcal{B} is the variety of its Borel subgroups, then $\text{Aut}^0(\mathcal{B}) = G$.

In conclusion, the classification of homogeneous complete varieties is equivalent to the classification of abelian varieties, semisimple groups of adjoint type (both of them well studied, see [8], [9], [6]) and parabolic subgroups P of a given semisimple group G (modulo automorphisms) such that $\text{Aut}^0(G/P) = G$. Therefore, the classification is essentially reduced to the classification of (non-exceptional, in the sense of Demazure [4]) parabolic subgroups of a simple group of adjoint type.

Finally, from the classification of the parabolic subgroups (in characteristic $p \neq 2, 3$) due to Wenzel ([12]) and the determination of the non-exceptional and reduced parabolic subgroups due to Demazure ([4]), we shall classify the parabolic varieties in characteristic $p \neq 2, 3$.

1. KNOWN RESULTS

The following results are well known and may be found in [5], [6], [8], [9], [11].

Theorem 1.1 (Barsotti-Chevalley). *Let \overline{G} be a smooth and connected algebraic group. There exists a unique normal subgroup $G \subset \overline{G}$ that is affine, smooth and connected, such that the quotient \overline{G}/G is an abelian variety.*

Definition 1.2. An affine algebraic group is said to be linearly reductive if any linear representation splits as a direct sum of simple representations.

An example of a linearly reductive group is the torus, and this is the only smooth and connected one in positive characteristic.

Theorem 1.3. *If a linearly reductive group acts on a smooth variety X , then the closed subscheme X^G of the fixed points is a smooth subscheme whose tangent space at each point x is the subspace of the G -invariant tangent vectors of X at x .*

Definition 1.4. A parabolic subgroup of an affine, smooth and connected group G is a subgroup $P \subset G$ such that the quotient variety G/P is complete.

Definition 1.5. A parabolic variety is a complete variety of the form G/P , where G is an affine, smooth and connected group and $P \subset G$ is a subgroup. A parabolic variety is called a flag variety if P is a reduced scheme.

Theorem 1.6 (Chevalley). *If G is an affine group and $H \subset G$ a subgroup, then G/H is a quasiprojective scheme and the action of G on G/H is projective, i.e., there exists a linear representation E of G such that G/H is a sub- G -scheme of $\mathbb{P}(E)$.*

Remark 1.7. It follows from this theorem that any parabolic variety is projective.

Theorem 1.8. *Any flag variety is a Fano variety; that is, the dual of the dualizing sheaf is ample.*

Theorem 1.9. *Let G be a semisimple group, $\text{Aut } G$ the automorphism group of G , and $I \subset \text{Aut } G$ the subgroup of the inner automorphisms of G . Then $(\text{Aut } G)/I$ is finite. In particular, if G is of adjoint type, then the connected component through the origin of $\text{Aut } G$ is isomorphic to G .*

Theorem 1.10. *Let X be an abelian variety, and $\text{Aut}^0 X \subset \text{Aut } X$ the connected component through the origin of the automorphism group of X (as a variety). The inclusion $X \hookrightarrow \text{Aut}^0 X$, $x \mapsto t_x = \text{translation by } x$, is an isomorphism.*

Theorem 1.11 (Borel's fixed point theorem). *If a smooth, solvable and connected affine group acts on a complete variety, then the subscheme of the fixed points is not empty.*

Definition 1.12. We shall denote by $\text{Pic}(X)$ the scheme in groups (if it exists) that parametrizes the invertible sheaves on X and by $\text{Pic}^0(X)$ the connected component of $\text{Pic}(X)$ through the origin.

Theorem 1.13. *If X is a complete variety, then $\text{Pic}(X)$ exists; it is a proper scheme and the tangent space at the origin is isomorphic to $H^1(X, \mathcal{O}_X)$. Moreover,*

- (1) *If X is an abelian variety, then $\text{Pic}(X)$ is smooth; hence, $\text{Pic}^0(X)$ is also an abelian variety.*
- (2) *If X is a parabolic variety, then $\text{Pic}^0(X) = 0$; in particular, $\text{Pic}(X)$ is a discrete k -scheme (i.e., $\coprod \text{Spec } k$). Furthermore, if $X = G/P$ where G is semisimple and simply connected, then $\text{Pic}(X)$ is isomorphic to the character group of P .*

Let X be a variety such that $\text{Pic}(X)$ exists. Let \mathcal{P} be a universal invertible sheaf on $X \times \text{Pic}^0(X)$, i.e., an invertible sheaf such that the pair $(\text{Pic}^0(X), \mathcal{P})$ represents the functor of invertible sheaves of degree zero on X . By the universal property, \mathcal{P} is univocally determined up to inverse images of invertible sheaves (of degree 0) on $\text{Pic}^0(X)$ by the projection $X \times \text{Pic}^0(X) \rightarrow \text{Pic}^0(X)$. Hence, if one fixes a point $x_0 \in X$ and requires that \mathcal{P} be trivial on $x_0 \times \text{Pic}^0(X)$, then \mathcal{P} is completely determined.

Given \mathcal{P} , one defines $\varphi: X \rightarrow \text{Pic}^0(\text{Pic}^0(X))$ by $\varphi(x) = \mathcal{P}|_{x \times \text{Pic}^0(X)}$. If, in order to determine \mathcal{P} , one replaces the chosen point x_0 by another one, x'_0 , then one obtains the same morphism composed with the translation on $\text{Pic}^0(\text{Pic}^0(X))$ by

the invertible sheaf $\varphi(x'_0)^{-1}$. That is, φ is univocally determined, up to translations on $\text{Pic}^0(\text{Pic}^0(X))$, by the choice of a point of X . If X is a group, one chooses as x_0 the origin of the group; this is equivalent to saying that φ is a morphism of groups ($\varphi(0) = 0$).

Theorem 1.14 (Duality of abelian varieties). *If A is an abelian variety, then the natural morphism $A \rightarrow \text{Pic}^0(\text{Pic}^0(A))$ is an isomorphism.*

Theorem 1.15. *If X, Y are complete varieties, then*

$$\text{Pic}^0(X) \times \text{Pic}^0(Y) = \text{Pic}^0(X \times Y),$$

and the correspondence is $(\mathcal{L}, \mathcal{L}') \mapsto \mathcal{L} \otimes_k \mathcal{L}'$.

2. LOCAL TRIVIALITY OF DEFORMATIONS OF PARABOLIC SUBGROUPS

Let us assume that k is an algebraically closed field of arbitrary characteristic p , and let G be an affine, smooth and connected algebraic group over k .

Definition 2.1. The nilpotence degree of an algebraic group G (over an algebraically closed field) is the order of the quotient by its reduced subgroup: $|G/G_{\text{red}}|$. That is, it is the dimension of the finite vector space of the functions of G/G_{red} .

It is known that a subgroup $P \subset G$ is parabolic if and only if it contains a Borel subgroup B of G . Moreover, given a Borel subgroup B , each parabolic subgroup has a conjugated one containing B , because all Borel subgroups are conjugate. Therefore, the problem is to determine the structure of the parabolic subgroups containing a given Borel subgroup.

The aim of this section is to prove the local triviality of any deformation of parabolic subgroups (containing a Borel subgroup B). More precisely, let us denote $X_S = X \times S$ for each k -scheme S . Then,

Theorem 2.2. *Let S be a connected scheme and $\tilde{P} \subset G_S$ a subscheme in groups over S containing B_S . Then:*

- (1) *If \tilde{P} is flat over S , then it is constant; that is, $\tilde{P} = P_S$ for some subgroup $P \subset G$ containing B .*
- (2) *If S is integral and the fibres of $\tilde{P} \rightarrow S$ have constant dimension and nilpotence degree, then \tilde{P} is constant: $\tilde{P} = P_S$.*

One first observes that any parabolic subgroup contains the radical $R(G)$ of the group (the maximal solvable, normal, smooth and connected subgroup of G). Hence, the parabolic subgroups are in biunivocal correspondence with the parabolic subgroups of the quotient $G/R(G)$, which is a semisimple group. Therefore, we shall assume in the following that G is semisimple.

Let $T \subset B$ be a maximal torus of G , \mathcal{R} the roots system of G associated with T , \mathcal{R}^+ the positive roots (i.e., the roots of (B, T)) and $\mathcal{S} \subset \mathcal{R}$ the basis of simple roots contained in \mathcal{R}^+ .

Remark 2.3. The reduced parabolic subgroups (containing B) are in biunivocal correspondence with the subsets I of the basis \mathcal{S} . In particular, the number of them is finite (exactly $2^{|\mathcal{S}|}$). Therefore, the theorem is almost immediate in characteristic 0, since any group is reduced, and hence the scheme of parabolic subgroups containing B is finite and discrete. Thus, for the proof of the theorem, we shall

assume that the characteristic is positive, $p > 0$, although the same proof is valid in characteristic 0 with the corresponding simplifications.

Let \mathcal{B} be the variety of Borel subgroups of G . It is known that $\mathcal{B} = G/B$, as G -varieties. Let $U \subset B$ be the unipotent part of B and U^- the unipotent part of the Borel subgroup B^- opposite to B (that is, the unique Borel subgroup containing T and such that $B \cap B^- = T$). One has that U^- is identified, as a U^- -scheme, with an open subset \mathcal{U}^- of \mathcal{B} (precisely, with the open set of the Borel subgroups B' such that $B' \cap U^- = \{e\}$), which coincides with the unique open orbit under the action of B^- .

Lemma 2.4. *Let $S \rightarrow \operatorname{Spec} k$ be a base change. The map $\tilde{P} \mapsto \tilde{P} \cap U_S^-$, between the subschemes in groups of G_S over S containing B_S and the subschemes in groups of U_S^- over S stable by T , is injective. Moreover, flat subschemes correspond to flat subschemes.*

Proof. Each subscheme in groups of G_S , $\tilde{P} \supset B_S$, defines a closed subscheme $\tilde{P}/B_S \subset (G/B)_S = \mathcal{B}_S$, which determines \tilde{P} , since \tilde{P} is the preimage of that closed subscheme by the morphism to the quotient $\pi: G_S \rightarrow G_S/B_S = \mathcal{B}_S$. Moreover, $U_S^- \cap (\tilde{P}/B_S)$ is dense in \tilde{P}/B_S (because it is dense in the fibres over S , since any parabolic subgroup is irreducible). Hence, \tilde{P}/B_S is the closure of $U_S^- \cap (\tilde{P}/B_S)$ in \mathcal{B}_S . It follows that the closed subscheme $U_S^- \cap (\tilde{P}/B_S) \subset U_S^-$ determines \tilde{P} . However, the identification of U^- with \mathcal{U}^- induces another one of U_S^- with \mathcal{U}_S^- , and it is clear that $U_S^- \cap (\tilde{P}/B_S)$ is identified with $U_S^- \cap \tilde{P}$, which is a subscheme in groups of U_S^- and is stable under the action of T by conjugation (since both subschemes are stable). In conclusion, $U_S^- \cap \tilde{P}$ determines \tilde{P} . The flatness of $\tilde{P} \cap U_S^-$ when \tilde{P} is flat follows from the flatness of $\tilde{P}/B_S \subset \mathcal{B}_S$ (this one is flat by descent, since $\tilde{P} \rightarrow \tilde{P}/B_S$ is faithfully flat), since $\tilde{P} \cap U_S^-$ is identified with the open subscheme $(\tilde{P}/B_S) \cap \mathcal{U}_S^-$. \square

In the case that $S = \operatorname{Spec} k$, this lemma may be obtained from Proposition 4 of [12] as a particular case. Therefore, this lemma generalizes that proposition for the relative case.

Let $\{\alpha_1, \dots, \alpha_s\} = R^-$ be the roots corresponding to B^- . By definition, these are the characters of T appearing in the action of T on the tangent space of U^- (so $s = \dim U^-$). They are nontrivial characters, and none of the α_i is a power of α_j for $i \neq j$ (in particular, they are different). Moreover, for each α_i there exists a unique additive subgroup $G_a^{\alpha_i} \subset U^-$ stable by T and such that T acts by multiplication by the character α_i , in such a way that the multiplication morphism $G_a^{\alpha_1} \times \dots \times G_a^{\alpha_s} \rightarrow U^-$ is an isomorphism of T -varieties.

If $\tilde{P} \subset G_S$ is a parabolic subgroup containing B , then, in particular, $\tilde{P} \cap U_S^- = H$ is a subgroup of U_S^- stable under the action of T by conjugation. Therefore, it suffices to study the structure of these subgroups.

Definition 2.5. A cone is the spectrum of an \mathbb{N} -graded algebra.

One observes that a \mathbb{Z} -graduation on an algebra A is equivalent to an action of algebras of the multiplicative group G_m on A .

Definition 2.6. We say that an action of G_m on $\text{Spec } A$ is conic if the corresponding graduation on A is an \mathbb{N} -graduation.

Proposition 2.7. If G_m acts on an affine scheme X in a conic way and $Y \subset X$ is a stable closed subscheme, then the action of G_m on Y is conic.

Proof. Immediate. \square

Remark 2.8. The action of the maximal torus T on U^- (with the above notation) is conic for certain one-parameter subgroups $G_m \hookrightarrow T$.

Lemma 2.9. The system of negative roots can be ordered in such a way that for each $i \leq s$ there exists a multiplicative subgroup $G_m^{(i)} \subset T$ such that the restriction of the root α_j to $G_m^{(i)}$ is a nontrivial negative character for $j > i$ and is the trivial character for $j = i$.

Proof. Using the additive notation for the characters of T , one has that $R^- \subset X(T) = \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$; thus, R^- are vectors with negative coordinates. One then has to find linear forms $\omega_i: X(T) \rightarrow \mathbb{Z}$ (which correspond to multiplicative subgroups of T) such that $\omega_i(\alpha_j)$ are negative for $j > i$ and $\omega_i(\alpha_i) = 0$, for a certain order of the indices. By recurrence, it clearly suffices to prove that for any subset $J = \{\beta_1, \dots, \beta_r\} \subset R^-$ there exists a reordering of the indices and a linear form $\omega: X(T) \rightarrow \mathbb{Z}$ such that ω is negative on β_j for $j > 1$ and $\omega(\beta_1) = 0$. Since R^- is contained in an open half-plane of the corresponding real vector space, any $J \subset R^-$ generates a convex cone that contains no full line. Let $\beta_1 \in J$ generate an extremal ray of that cone. Then there exist a linear form ω vanishing at β_1 , and strictly negative on all other $\beta \in J$ (since no negative roots are proportional); we may further assume that ω is rational, and even integral. \square

Corollary 2.10. Let us fix an ordering of the roots given by the preceding proposition. If we denote $U_i^- = G_a^{\alpha_i} \times \cdots \times G_a^{\alpha_s}$, then

- U_i^- is closed in U^- and stable under the action of T ;
- $G_m^{(i)}$ acts in a conic way on U_i^- , and the subscheme of the fixed points is $(U_i^-)^{G_m^{(i)}} = G_a^{\alpha_i}$.

Proposition 2.11.

(1) Let H be an affine S -scheme in groups on which G_m acts in a conic way through automorphisms of S -schemes in groups. Then, there exists a unique retract $\rho: H \rightarrow H^{G_m}$ of G_m -schemes, where H^{G_m} is the closed subscheme in groups of the fixed points of H by G_m . Moreover, ρ is a morphism of S -schemes in groups, and hence $\ker \rho = N$ is the unique subscheme in groups of H stable by G_m such that

$$H = N \rtimes_S H^{G_m}.$$

(2) If H is an affine S -scheme in groups on which G_m acts in a conic way through automorphisms of S -schemes in groups and $\overline{H} \subset H$ is a subscheme in groups stable by G_m , then G_m acts in a conic way on \overline{H} and the restriction of the retract $\rho: H \rightarrow H^{G_m}$ to \overline{H} is the corresponding retract $\rho_{\overline{H}}: \overline{H} \rightarrow \overline{H}^{G_m} \subset H^{G_m}$.

Proof. (1) By the uniqueness of ρ it suffices to prove the statement locally on S . That is, we may assume $S = \text{Spec } k$ with k a ring, and $H = \text{Spec } A$, with A a Hopf k -algebra.

By hypothesis, A is a Hopf T -algebra, i.e., an \mathbb{N} -graded Hopf k -algebra. Let us consider the Hopf subalgebra $A_0 \subset A$ of the elements of degree zero. Let $I \subset A$ be the irrelevant ideal. One has that I is the ideal of a subgroup of H over k and $A_0 \stackrel{\pi}{\simeq} A/I$ as Hopf k -algebras. This means that the inclusion $\text{Spec } A/I \subset \text{Spec } A$ has the retract $\pi_0: \text{Spec } A \rightarrow \text{Spec } A_0$. Now, $\text{Spec } A/I = H^{G_m}$, since a point of H valued at B , $h: A \rightarrow B$ (morphism of k -algebras), is a fixed point if and only if h is a morphism of G_m -algebras, where B is endowed with the trivial structure of G_m -algebra; in other words, if and only if it is a morphism of graded k -algebras with the trivial graduation on B ($B = B_0$). But this is equivalent to saying that $h(I) = 0$, i.e., h factors through A/I or h is a point of $\text{Spec } A/I$ valued at B . This yields the existence of ρ and proves that it is a morphism of schemes in groups.

The uniqueness is given from the fact that if a morphism $\rho: A/I \rightarrow A$ is a section of algebras and G_m -algebras of the projection $\pi: A \rightarrow A/I$, then it maps into A_0 (since it is a morphism of G_m -algebras) and, since it is a section of π , it must be the inverse of the isomorphism $\pi|_{A_0}: A_0 \rightarrow A/I$.

(2) It is immediate that the action is conic. For the second part, in the above reduction, the retract is given by the inclusion $A_0 \subset A$; if A is the ring of H and \overline{A} the ring of \overline{H} , then the statement follows from the commutativity of the diagram

$$\begin{array}{ccc} A_0 & \hookrightarrow & A \\ \pi \downarrow & & \downarrow \pi \\ \overline{A}_0 & \hookrightarrow & \overline{A} \end{array}$$

□

Corollary 2.12. *With the notation of Corollary 2.10, one has that U_i^- is a subgroup of U^- , and it is the kernel of the unique retract of groups and T -schemes $\rho_{i-1}: U_{i-1}^- \rightarrow G_a^{\alpha_{i-1}}$.*

Proof. (By recurrence on i). One has that U_{i-1}^- is a subgroup of U^- . Let us consider the subgroup $G_m^{(i-1)} \subset T$ of Lemma 2.9. From Corollary 2.10 one has that $(U_{i-1}^-)^{G_m^{(i-1)}} = G_a^{\alpha_{i-1}}$ and, by Proposition 2.11, there exists a unique retract $\rho_{i-1}: U_{i-1}^- \rightarrow G_a^{\alpha_{i-1}}$ of $G_m^{(i-1)}$ -schemes which is, in addition, a morphism of groups. Therefore, it must coincide with the projection on $G_a^{\alpha_{i-1}}$, and hence $\ker \rho_{i-1} = U_i^-$; in particular, U_i^- is a subgroup of U^- and ρ_{i-1} is a morphism of T -schemes. □

Proposition 2.13. *The subschemes in groups H of U_S^- stable by T are subschemes of the form*

$$H = H_1 \times_S \cdots \times_S H_s \subset U_S^-$$

with $H_i \subset (G_a^{\alpha_i})_S$ a subscheme in groups stable by homotheties.

Proof. In order to simplify the notation, let us assume that $S = \text{Spec } k$, i.e., $H \subset U^-$, since the proof does not depend on the base S . Let us denote $T_1 = G_m^{(1)} \subset T$. By Corollary 2.10, U^- and H are cones with respect to T_1 , and the fixed points are respectively $(U^-)^{T_1} = G_a^{\alpha_1}$ and $H^{T_1} = H \cap (U^-)^{T_1} = H \cap G_a^{\alpha_1} = H_1$. Moreover, the retract $\rho: U^- \rightarrow G_a^{\alpha_1}$ of T_1 -schemes in groups restricts to the corresponding retract $\rho|_H: H \rightarrow H_1$ (Proposition 2.7). Thus, since $\ker \rho = U_2^-$, one concludes that $H = H_1 \times \overline{H}$ with $H_1 \subset G_a^{\alpha_1}$ and $\overline{H} \subset U_2^-$. Recurrently, one concludes. □

Proposition 2.14.

(1) If S is a connected scheme, then the subschemes in groups of U_S^- flat over S and stable by T are constant; that is, they are all obtained from the subgroups of U^- by the base change $S \rightarrow \operatorname{Spec} k$.

(2) If S is integral, then the subschemes in groups of U_S^- stable by T and flat over S are precisely the ones with constant dimension and nilpotence degree along the fibres over S .

Proof. We may assume $S = \operatorname{Spec} A$. By the preceding proposition, any subgroup H of U_S^- stable by T (without flatness conditions over S) has the form $H_1 \times_S \cdots \times_S H_s$, with $H_i \subset (G_a^{\alpha_i})_S$ a stable subgroup by T . We may therefore assume $U^- = G_a = \operatorname{Spec} k[Y]$ and $H \subset (G_a)_S$ stable by homotheties. In this case, $H = \operatorname{Spec} \bigoplus_i A/I_i \cdot Y^i$ for certain ideals $I_i \subset A$ satisfying $I_0 \subset I_1 \subset \cdots \subset I_n \subset \cdots$.

(1) If H is flat over S , then the A -modules A/I_i are also flat; thus, they are locally free with rank less than or equal to 1, i.e., either $I_i = 0$ or $I_i = A$. It is then clear that H descends to U^- .

(2) Assume S is integral. If $H \subset (G_a)_S$ has constant dimension 1 along the fibres, then $H = (G_a)_S$, since the equality holds in the fibres over S . Therefore, $(I_i)_0 = S$ and hence $I_i = 0$ (since S is integral). If H has constant dimension 0 along the fibres, then for each point $x \in S$ the nilpotence degree is the maximum index i such that $\mathfrak{p}_x \supset I_i$ ($\mathfrak{p}_x \subset A$ being the prime ideal of the functions that are zero at x). Since the nilpotence degree is constant along the fibres, one has that either $(I_i)_0 = S$ or $(I_i)_0 = \emptyset$; that is, either $I_i = 0$ or $I_i = A$ for each i , and one concludes. \square

Proof of Theorem 2.2. This follows from Lemma 2.4 and Proposition 2.14. \square

3. TRIVIALITY OF THE ACTIONS WITH COMPLETE ORBITS

Definition 3.1. Let X be a variety on which an algebraic group G acts and let $x \in X$. We shall call the nilpotence degree of the orbit of x the nilpotence degree of its isotropy group: $|P_x/(P_x)_{\text{red}}|$.

Remark 3.2. In characteristic zero this notion is superfluous, since any group is smooth and hence the nilpotence degree of any orbit is 1.

Theorem 3.3. Let G be an affine, smooth and connected algebraic group acting on a smooth variety X . Assume that the orbits of the action are complete and that they have constant dimension and nilpotence degree. Then all the orbits are isomorphic as G -varieties to a given one Y and $X \simeq Y \times X/G$ as G -varieties, where the action of G on X/G is trivial. In particular, X/G is a geometric quotient of X by G .

Proof. Choose B , a Borel subgroup. Let X^B be the reduced subscheme of the fixed points of X under the action of B . Then B fixes a unique point in any G -orbit in X . So the map $G \times X^B \rightarrow X$ is surjective and its fiber at $x \in X^B$ is G_x (i.e., the isotropy group of x), which is irreducible and of constant dimension (since the dimension of G/G_x is constant). Thus, X^B is irreducible.

Let us define $\phi: G \times X^B \rightarrow X \times X^B$ by $\phi(g, x) = (g \cdot x, x)$ and let $\tilde{P} = \phi^{-1}(\Delta_{X^B}) \subset G \times X^B$, where $\Delta_{X^B} = X^B \subset X \times X^B$ is the diagonal. It is a subscheme in groups of $G \times X^B$ over X^B ; it contains $B \times X^B$, and its fibre in each

point $x \in X^B$ is the isotropy group of x that, by hypothesis, has constant dimension and nilpotence degree. Since X^B is connected and integral, one concludes, by Theorem 2.2, that it is constant: $\tilde{P} = \bar{P} \times X^B$, with $\bar{P} \subset G$ a subgroup containing B . It thus follows that φ factors through an isomorphism $G/\bar{P} \times X^B \rightarrow X$. \square

4. INVARIANT INVERTIBLE SHEAVES

Definition 4.1. Let X be a complete and homogeneous variety. An invertible sheaf $\mathcal{L} \in \text{Pic}(X)$ is said to be invariant if it is a fixed point of $\text{Pic}(X)$ under the natural action of $\text{Aut}^0(X)$ in $\text{Pic}(X)$.

Remark 4.2. If X is a parabolic variety, then any invertible sheaf is invariant, since $\text{Pic}_{\text{red}}(X)$ is a discrete k -scheme (because $\text{Pic}^0(X) = 0$) and $\text{Aut}^0(X)$ is smooth and connected.

Proposition 4.3. Let X be a complete and homogeneous variety. If $\mathcal{L} \in \text{Pic}(X)$ is effective and invariant, then the complete linear system $\Gamma(X, \mathcal{L})$ has no base points, $\text{Aut}^0(X)$ acts by automorphisms on $\mathbb{P}(H^0(X, \mathcal{L})^*)$, and the natural morphism $\phi_{\mathcal{L}}: X \rightarrow \mathbb{P}(H^0(X, \mathcal{L})^*)$ is an $\text{Aut}^0(X)$ -morphism.

Proof. Let $G = \text{Aut}^0(X)$, and let $m: G \times X \rightarrow X$ be the action. Since \mathcal{L} is invariant, one has $m^*\mathcal{L} \simeq \mathcal{L}_G \otimes_k \mathcal{L}$, for some invertible sheaf \mathcal{L}_G on G . Taking direct images over G , one obtains $1 \otimes m^*: \mathcal{O}_G \otimes_k H^0(X, \mathcal{L}) \xrightarrow{\sim} \mathcal{L}_G \otimes_k H^0(X, \mathcal{L})$, and hence a morphism of G -schemes

$$G \times \mathbb{P}(H^0(\mathcal{L})) = \mathbb{P}_G(\mathcal{O}_G \otimes_k H^0(\mathcal{L})) \xrightarrow{1 \otimes m^*} \mathbb{P}_G(\mathcal{L}_G \otimes_k H^0(\mathcal{L})) = G \times \mathbb{P}(H^0(\mathcal{L})),$$

where we have denoted $H^0(\mathcal{L}) = H^0(X, \mathcal{L})$. Taking a fibre at each point of G , one obtains a morphism $\tau: G \rightarrow \text{Aut}_k(\mathbb{P}(H^0(X, \mathcal{L}))) = \mathbf{PGL}_k(H^0(X, \mathcal{L}))$ such that for any rational point $g \in G$ and any section $s \in H^0(X, \mathcal{L})$ one has $\tau_g(\langle s \rangle) = \langle g^*s \rangle \subset H^0(X, g^*\mathcal{L}) = H^0(X, \mathcal{L})$. Hence the transposed morphism $\bar{\tau}: G \rightarrow \mathbf{PGL}_k(H^0(X, \mathcal{L})^*)$ satisfies $\bar{\tau}_g(\langle \omega \rangle) = \langle g \cdot \omega \rangle$, for any $g \in G$ and $\omega \in H^0(X, \mathcal{L})^*$, where $(g \cdot \omega)(s) = \omega(g^*s)$ for each $s \in H^0(X, \mathcal{L})$. It follows easily that $\bar{\tau}$ is a morphism of groups (notice that a morphism between two smooth groups is a morphism of groups if and only if it is so for the rational points). Moreover, if for each point $x \in X$ we denote by $\omega_x: H^0(X, \mathcal{L}) \rightarrow \mathcal{L}_x \simeq k$ the linear form that maps each section to its fibre at x , then $\langle g \cdot \omega_x \rangle = \langle \omega_{g \cdot x} \rangle$. Therefore, the closed subset of base points is empty, since it is G -invariant, \mathcal{L} is effective, and X is homogeneous, and the natural morphism $\phi_{\mathcal{L}}: X \rightarrow \mathbb{P}(H^0(X, \mathcal{L})^*)$, defined by $\phi_{\mathcal{L}}(x) = \langle \omega_x \rangle$, is a G -morphism. \square

Remark 4.4. This theorem implies the following well-known results:

- (1) If X is an abelian variety, $\mathcal{L} \in \text{Pic}^0(X)$ is nontrivial if and only if $H^0(X, \mathcal{L}) = 0$.
- (2) A complete and homogeneous variety X is parabolic if and only if $\text{Aut}^0(X)$ is an affine group (and, in that case, it is semisimple of adjoint type).

Definition 4.5. We shall say that a morphism $\pi: X \rightarrow Y$ of smooth varieties is a quotient of X if

- (1) π is surjective (as a map),
- (2) π is submersive, i.e., a subset $U \subset Y$ is open if and only if $\pi^{-1}(U)$ is open in X ,

(3) $\mathcal{O}_Y = \ker(Id \otimes 1 - 1 \otimes Id)$, with $Id \otimes 1 - 1 \otimes Id: \pi_*\mathcal{O}_X \rightarrow \pi_*\mathcal{O}_X \otimes_{\mathcal{O}_Y} \pi_*\mathcal{O}_X$ the natural morphism.

We shall say that a quotient $\pi: X \rightarrow Y$ is an invariant quotient if

(4) there exists an action of $\text{Aut}^0(X)$ on Y such that π is a $\text{Aut}^0(X)$ -morphism.

This notion of quotient corresponds to the notion of quotient by an equivalence relation in the theory of schemes; that is, it is wider than Mumford’s notion of geometric quotient under the action of a group.

Remark 4.6. If X is a homogeneous variety, then one morphism $\pi: X \rightarrow Y$ (where Y is a smooth variety) is an invariant quotient of X if and only if it is surjective and invariant (i.e., it satisfies the conditions (1) and (4)). Indeed, if π is invariant, then the fibers are equidimensional and so π is flat (X, Y are smooth varieties). Moreover, if π is surjective, then π is faithful flat, and one concludes easily, by flat descent, that π satisfies conditions (2) and (3).

Definition 4.7. We shall say that two quotients $\pi: X \rightarrow Y, \pi': X \rightarrow Y'$ are equivalent if there exists an isomorphism $\phi: Y \rightarrow Y'$ such that $\pi' = \phi \circ \pi$.

Remark 4.8. Let X be a homogeneous variety. An invariant quotient $\pi: X \rightarrow Y$ is determined, up to equivalence, by the fibre passing through a given point $x \in X$, that is, by the subscheme $\pi^{-1}(\pi(x)) \subset X$. Indeed, if $G = \text{Aut}^0(X)$ and H is the isotropy group of x , then $X = G/H, Y = G/H'$ (with H' the isotropy group of $\pi(x)$) and $\pi^{-1}(\pi(x)) = H'/H \subset G/H$.

Remark 4.9. The invariant quotients of X , up to equivalence, are partially ordered in the following way: $(Y, \pi) \geq (Y', \pi')$ if there exists a morphism $h: Y \rightarrow Y'$ of $\text{Aut}^0(X)$ -varieties such that $\pi' = h \circ \pi$. Moreover, they form an inverse system: if $(Y_1, \pi_1), (Y_2, \pi_2)$ are two invariant quotients, there exists a third one, (Y, π) , such that $(Y, \pi) \geq (Y_1, \pi_1), (Y_2, \pi_2)$; it is enough to define Y as the image of the $\text{Aut}^0(X)$ -morphism $\pi_1 \times \pi_2: X \rightarrow Y_1 \times Y_2$. Notice that if Y_1 and Y_2 are parabolic varieties (respectively, abelian varieties), then Y is parabolic (respectively, abelian).

Lemma 4.10. *Let X be a complete variety, \mathcal{L} an invertible sheaf on X and $V_2 \subset V_1 \subset H^0(X, \mathcal{L})$ two linear systems without base points. Let $\pi_i: X \rightarrow X_i \subset \mathbb{P}(V_i^*)$ be the morphism induced by V_i , with $X_i = \text{Im } \pi_i$. There exists a morphism $h: X_1 \rightarrow X_2$ such that $\pi_2 = h \circ \pi_1$. Moreover, h is finite.*

Proof. Let $(V_2)_0 \subset V_1^*$ be the subspace incident with V_2 . Let us consider the natural projection $\bar{h}: \mathbb{P}(V_1^*) - \mathbb{P}((V_2)_0) \rightarrow \mathbb{P}(V_2^*)$. Then $\pi_1^{-1}(X_1 \cap \mathbb{P}((V_2)_0))$ is the subscheme of base points of V_2 , which is empty; hence \bar{h} induces a morphism $h: X_1 \rightarrow X_2$.

The finiteness of h follows from the fact that it is an affine morphism, since \bar{h} is affine. □

Proposition 4.11. *Let X be a complete and homogeneous variety. There exists a parabolic (respectively, abelian) invariant quotient $\pi: X \rightarrow \mathcal{P}(X)$ satisfying the following universal property: for any parabolic (respectively, abelian) invariant quotient $\bar{\pi}: X \rightarrow \bar{\mathcal{P}}$, there exists a unique morphism $f: \mathcal{P}(X) \rightarrow \bar{\mathcal{P}}$ such that $\bar{\pi} = f \circ \pi$. Moreover, there exists an effective and invariant invertible sheaf \mathcal{L} on X such that $\mathcal{P}(X) = \text{Proj } \bigoplus_{n \in \mathbb{N}} H^0(X, \mathcal{L}^n)$ and the natural morphism*

$$X \longrightarrow \text{Proj } \bigoplus_{n \in \mathbb{N}} H^0(X, \mathcal{L}^n) = \mathcal{P}(X)$$

coincides with π .

Proof. The invariant quotients are determined by the fibre through a given point x , in such a way that the order of the quotients corresponds with the inclusion order of the fibres; that is, $(Y, \pi) \geq (\bar{Y}, \bar{\pi})$ if and only if $\pi^{-1}(\pi(x)) \subset \bar{\pi}^{-1}(\bar{\pi}(x))$. Since X is Noetherian, one concludes the existence of a maximal parabolic (respectively, abelian) quotient.

For the second part, let $\bar{\mathcal{L}}$ be an ample invertible sheaf on $\mathcal{P}(X)$. Since $\mathcal{P}(X)$ is parabolic, $\bar{\mathcal{L}}$ is invariant, and so $\mathcal{L} = \pi^*\bar{\mathcal{L}}$ is invariant too. Let us consider the parabolic quotient $\bar{\pi}: X \rightarrow \bar{\mathcal{P}} = \text{Proj} \bigoplus_{n \in \mathbb{N}} H^0(X, \mathcal{L}^n)$. One has that $\mathcal{P}(X) = \text{Proj} \bigoplus_{n \in \mathbb{N}} H^0(\mathcal{P}(X), \bar{\mathcal{L}}^n)$ and $H^0(\mathcal{P}(X), \bar{\mathcal{L}}^n) \subset H^0(X, \mathcal{L}^n)$. This defines a morphism $h: \bar{\mathcal{P}} \rightarrow \mathcal{P}(X)$ such that $\pi = h \circ \bar{\pi}$ (Lemma 4.10). Hence $(\bar{\mathcal{P}}, \bar{\pi}) \geq (\mathcal{P}(X), \pi)$ and, by maximality, one concludes that $(\bar{\mathcal{P}}, \bar{\pi}) = (\mathcal{P}(X), \pi)$. \square

Definition 4.12. The universal parabolic quotient $\pi: X \rightarrow \mathcal{P}(X)$ of the latter proposition is called the parabolic part of X . The universal abelian quotient $\pi_{ab}: X \rightarrow \text{Ab}(X)$ is called the abelian part of X .

5. STRUCTURE OF COMPLETE AND HOMOGENEOUS VARIETIES

Lemma 5.1. *Let G be a smooth and connected group. If G is a normal subgroup of a smooth and connected group \bar{G} and \bar{G} acts transitively on a variety Y , then the orbits of Y under the action of G are closed and conjugated by \bar{G} (in particular, the orbits are isomorphic as G -schemes). In addition, there exists a fine quotient $Y \rightarrow Y/G$.*

Proof. One has that $Y \simeq \bar{G}/H$, as \bar{G} -schemes. Since G is normal in \bar{G} , it follows that $G \cdot H \subset \bar{G}$ is a subgroup and it is clear that $Y/G \simeq \bar{G}/G \cdot H$ as \bar{G} -schemes, in such a way that the morphism $\pi: Y \rightarrow Y/G$ corresponds to $\bar{G}/H \rightarrow \bar{G}/G \cdot H$. One concludes easily. \square

Theorem 5.2. *If X is a complete and homogeneous variety over k , then:*

(1) *X splits, in a unique way, as the direct product of a parabolic variety and an abelian variety, $X = Y \times A$. These factors are canonically determined from X in the following way: A is the Albanese variety of X (that is, $A = \text{Pic}^0(\text{Pic}^0(X))$, where Pic^0 denotes the reduced connected component through the origin of the Picard scheme) and $\pi_1: X \rightarrow Y$ is the parabolic part of X .*

(2) *X is projective, and (the connected and reduced component of) its automorphism scheme splits, in a unique way, as a direct product of an abelian variety and a semisimple group of adjoint type. More precisely: if $X = Y \times A$, with Y the parabolic part and A the abelian one, then $\text{Aut}^0(X) = \text{Aut}^0(Y) \times A$ and $\text{Aut}^0(Y)$ is a semisimple group of adjoint type.*

Proof. Let $\bar{G} = \text{Aut}^0(X)$. Let $G \subset \bar{G}$ be the unique affine, smooth, connected and normal subgroup such that $A = \bar{G}/G$ is an abelian variety (Theorem 1.1). By Lemma 5.1, G acts on X with closed and isomorphic (as G -schemes) orbits. Then, by Theorem 3.3, $X = G/P \times X/G$, and $A = X/G$ is a quotient of $\bar{G}/G = A$ and hence an abelian variety.

For the uniqueness of the decomposition $X = Y \times A$, it is enough to determine canonically the projections of X onto the factors:

(1) Let Pic^0 be the reduced connected component through the origin of the Picard scheme of X , which is an abelian variety (Theorem 1.13(1)). It is known

that $Pic^0(Y) = 0$ (Theorem 1.13(2)) and that $Pic^0(Pic^0(A)) = A$ (Theorem 1.14), and this equality is canonical up to translations of A (that is, it is canonical once the origin is fixed). Moreover, $Pic^0(Y \times A) = Pic^0(Y) \oplus Pic^0(A)$ (Theorem 1.15), and one thus has a natural morphism

$$X \rightarrow Pic^0(Pic^0(X)) = Pic^0(Pic^0(Y) \times Pic^0(A)) = Pic^0(Pic^0(A)) = A,$$

which is precisely the projection. In other words, A is the Albanese variety of X , and the projection $X \rightarrow A$ is the natural morphism $X \rightarrow Alb(X)$.

(2) In order to see that $\pi_1 : X \rightarrow Y$ is the parabolic part of X , it is enough to see that $(Y, \pi_1) \geq (\mathcal{P}(X), \pi)$, i.e., that the fibre of a point $y_0 \in Y$ by π_1 , $y_0 \times A \subset X$, is contained in some fibre of π (Remark 4.8). By Proposition 4.11, $\pi : X \rightarrow \mathcal{P}(X)$ is the morphism induced by an effective invariant invertible sheaf \mathcal{L} on X ; hence it suffices to prove that \mathcal{L} is trivial on $y_0 \times A$. But, since $A = y_0 \times A$ is an abelian variety, any effective invariant invertible sheaf is trivial (Remark 4.4 (1)). The restriction of \mathcal{L} to A is effective because $H^0(X, \mathcal{L}) \neq 0$ has no base points (Proposition 4.3).

For the second part of the statement, it is enough to observe that, since the projections onto the parabolic and abelian parts are invariant quotients, any automorphism induces an automorphism of each factor, and conversely. The rest is immediate. □

Corollary 5.3. *Any group extension of an abelian variety by a semisimple group of adjoint type is trivial.*

Proof. Let A be an abelian variety, G a semisimple group of adjoint type, and let

$$0 \rightarrow G \rightarrow \overline{G} \rightarrow A \rightarrow 0$$

be a group extension. Let $B \subset G$ be a Borel subgroup of G and $X = \overline{G}/B$. Since G is a normal subgroup of \overline{G} , it follows easily that the orbits of X under the action of G are all isomorphic to G/B . Hence, by Theorem 3.3, $X = G/B \times A$ (since $\overline{G}/G = A$) and $\overline{G} \subset Aut^0(X) = Aut^0(G/B) \times Aut^0(A) = Aut^0(G/B) \times A$. Let $\pi_1 : Aut^0(G/B) \times A \rightarrow Aut^0(G/B)$ be the natural projection. Then $\pi_1(G) = Aut^0(G/B) = G$ (Demazure [4]). It follows that $\pi_1 : \overline{G} \rightarrow G$ is a retract of groups, and hence \overline{G} is the trivial extension. □

Definition 5.4. The parabolic type of a homogeneous variety X is the class, modulo automorphisms of $Aut^0(X)$, of the isotropy group of any point of the variety under the action of its group of automorphisms (more properly, the reduced connected component of its group of automorphisms).

Remark 5.5. From the preceding theorem, the isotropy group of any point x of a complete homogeneous variety X is a parabolic subgroup of the affine part of its group of automorphisms. Therefore, the parabolic type of the variety coincides with the class of its parabolic part modulo isomorphisms.

Theorem 5.6. *Two complete and homogeneous varieties are isomorphic if and only if their groups of automorphisms are isomorphic and they have the same parabolic type. In particular, the group of automorphisms classifies the variety, once its parabolic type has been given. In characteristic zero, there exist at most a finite number of complete homogeneous varieties with a given group of automorphisms.*

Given a complete and homogeneous variety $X = Y \times A$, let us denote by $\text{Aut}^{\text{lin}}(X)$ the maximum affine, normal, smooth and connected algebraic subgroup of $\text{Aut}^0(X)$ (as usual, $\text{Aut}^0(X)$ denotes the reduced connected component of the group of automorphisms of X), and let $\text{Aut}^{\text{Ab}}(X) = \text{Aut}^0(X)/\text{Aut}^{\text{lin}}(X) = \text{Alb}(X)$.

Theorem 5.7 (Classification of complete homogeneous varieties). *The classification of complete homogeneous varieties is equivalent to the classification of the triplets $(A, G, [P])$, where A is an abelian variety, G is a semisimple group of adjoint type, and $[P]$ is a parabolic type of G such that $G = \text{Aut}^0(G/P)$. The correspondence is*

$$\begin{aligned} X &\mapsto (\text{Alb}(X), \text{Aut}^{\text{lin}}(X), [P_x]), \\ A \times G/P &\leftarrow (A, G, [P]), \end{aligned}$$

where $P_x \subset \text{Aut}^{\text{lin}}(X)$ is the isotropy group of any point $x \in X$ and $[P_x]$ denotes its class modulo automorphisms of $\text{Aut}^{\text{lin}}(X)$.

6. CLASSIFICATION OF PARABOLIC VARIETIES

To conclude, we are going to give the classification of the parabolic varieties from the following results:

- The classification of the parabolic subgroups given by Wenzel in [12] (in characteristic different from 2 and 3).
- The determination (given by Demazure in [4]) of the pairs $P \subset G$, where G is a simple group of adjoint type and $P \subset G$ is a reduced parabolic subgroup such that $G = \text{Aut}^0(G/P)$. The pairs satisfying this condition are called non-exceptional. Demazure proves that the exceptional pairs are the following ones:

- (1) $G = \mathbf{SO}_{2l+1}(k)$ and G/P the variety that parameterizes the totally isotropic subspaces $V_l \subset k^{2l+1}$ (with $2l+1 \geq 5$). In this case $\text{Aut}^0(G/P) \simeq \mathbf{PSO}_{2l+2}$.
- (2) $G = \mathbf{Sp}_{2l}(k)$ and $G/P = \mathbb{P}_{2l-1}$ the variety that parameterizes the lines of k^{2l} . In this case $\text{Aut}^0(\mathbb{P}_{2l-1}) \simeq \mathbf{PGL}_{2l}(k)$.
- (3) G = the simple group of adjoint type with semisimple rank 2 and type G_2 ; that is, it is the group of automorphisms of an algebra of octonions Ω . Let $\tilde{\Omega} \subset \Omega$ be the hyperplane of the pure octonions and G/P the variety of isotropic lines of $\tilde{\Omega}$. Then G/P is isomorphic to a projective quadric of dimension 5, and hence, $\text{Aut}^0(G/P) = \mathbf{PSO}_6(k)$.

Theorem 6.1. *A complete homogeneous variety X splits as a direct product of two varieties $X = Y_1 \times Y_2$ if and only if the group of automorphisms $\text{Aut}^0(X)$ splits as a direct product of two groups $\text{Aut}^0(X) = G_1 \times G_2$. Moreover, one has:*

- (1) Y_1, Y_2 are complete and homogeneous varieties.
- (2) $\mathcal{P}(X) = \mathcal{P}(Y_1) \times \mathcal{P}(Y_2)$.
- (3) $\text{Ab}(X) = \text{Ab}(Y_1) \times \text{Ab}(Y_2)$.
- (4) $\text{Aut}^0(X) = \text{Aut}^0(Y_1) \times \text{Aut}^0(Y_2)$.

Proof. Assume that $X = Y_1 \times Y_2$. Since $Y_1 \simeq Y_1 \times y_2 \hookrightarrow X$ for any closed point $y_2 \in Y_2$, Y_1 is projective. Analogously, Y_2 is projective. Moreover, they are both reduced and connected, since the product $Y_1 \times Y_2 = X$ is so. In particular, $\text{Pic}^0(X) = \text{Pic}^0(Y_1) \times \text{Pic}^0(Y_2)$. It follows that $\text{Ab}(X) = \text{Ab}(Y_1) \times \text{Ab}(Y_2)$ and the morphism $\mu: X \rightarrow \text{Ab}(X)$ is precisely $\mu = \mu_1 \times \mu_2: X = Y_1 \times Y_2 \rightarrow \text{Ab}(Y_1) \times \text{Ab}(Y_2)$. On

the other hand, the fibre by μ of any point $x = (y_1, y_2)$ is $\mathcal{P}(X) = F_1 \times F_2$, where $F_1 \subset Y_1$, $F_2 \subset Y_2$ are the fibres by μ_1 and μ_2 of the points y_1 and y_2 , respectively.

Analogously, one proves that F_1, F_2 are projective and connected varieties. Let $\bar{\mathcal{L}}_1$ be an effective ample invertible sheaf on F_1 . Then $\mathcal{L}_1 = \bar{\mathcal{L}}_1 \otimes_k \mathcal{O}_{F_2}$ is an invertible sheaf on $\mathcal{P}(X)$, and hence invariant (since $\mathcal{P}(X)$ is a parabolic variety). The natural morphism

$$\mathcal{P}(X) \rightarrow \text{Proj} \bigoplus_{n \in \mathbb{N}} H^0(\mathcal{P}(X), \mathcal{L}_1^n) = \text{Proj} \bigoplus_{n \in \mathbb{N}} H^0(F_1, \bar{\mathcal{L}}_1^n) = F_1$$

is an invariant quotient, and it coincides with the projection on the first factor. One concludes that F_1, F_2 are $\text{Aut}^0(\mathcal{P}(X))$ -varieties and $\text{Aut}^0(\mathcal{P}(X)) = \text{Aut}^0(F_1) \times \text{Aut}^0(F_2)$. In particular, one has that

$$\begin{aligned} \text{Aut}^0(X) &= \text{Aut}^0(\mathcal{P}(X)) \times \text{Ab}(X) \\ &= (\text{Aut}^0(F_1) \times \text{Ab}(Y_1)) \times (\text{Aut}^0(F_2) \times \text{Ab}(Y_2)) = G_1 \times G_2; \end{aligned}$$

that is, $\text{Aut}^0(X)$ splits as a direct product of groups in such a way that G_1 acts trivially on Y_2 and G_2 acts trivially on Y_1 . So $\text{Aut}^0(X) = G_1 \times G_2 \subset \text{Aut}^0(Y_1) \times \text{Aut}^0(Y_2) \subset \text{Aut}^0(X)$, and then $G_1 = \text{Aut}^0(Y_1)$ and $G_2 = \text{Aut}^0(Y_2)$.

Conversely, if $\text{Aut}^0(X) = G_1 \times G_2$, then $\text{Aut}^{lin}(X) = G_1^{lin} \times G_2^{lin}$ and $\text{Aut}^{Ab}(X) = G_1^{Ab} \times G_2^{Ab}$. In particular, G_1^{lin} is a normal subgroup of $\text{Aut}^0(X)$, and this is an affine smooth and connected group. By Lemma 5.1, $X = F_1 \times X/G_1^{lin}$, and the projections are $\text{Aut}^0(X)$ -morphisms. F_1 is an orbit under the action of G_1^{lin} ; hence G_1^{lin} is semisimple of adjoint type, $F_1 = G_1^{lin}/P_1$ and the projection $X \rightarrow F_1$ is an $\text{Aut}^0(X)$ -morphism. Therefore, $G_1^{Ab} \times G_2$ induces a group of automorphisms of F_1 that commute with G_1^{lin} , i.e., a subgroup of $\text{Aut}_{G_1^{lin}\text{-var}}(G_1^{lin}/P_1)_{red} = (N_{G_1^{lin}}(P_1)/P_1)_{red} = 0$, where $N_{G_1^{lin}}(P_1)$ is the normalizer of P_1 in G_1^{lin} , because if $g \in G_1^{lin}$ normalizes P_1 , then it normalizes $(P_1)_{red}$, and hence $g \in (P_1)_{red} \subset P_1$, since $N_G(P) = P$ for any reduced parabolic subgroup P of an affine smooth and connected group G . Consequently, $\text{Aut}^0(F_1) = G_1^{lin}$ and $\text{Aut}^0(X/G_1^{lin}) = G_1^{Ab} \times G_2$. Repeating the same argument with G_2^{lin} and X/G_1^{lin} , one obtains that $X/G_1^{lin} = F_2 \times X/(G_1^{lin} \times G_2^{lin}) = F_2 \times \text{Ab}(X)$, where the projections are invariant quotients, $\text{Aut}^0(X/G_1^{lin}) = \text{Aut}^0(F_2) \times \text{Aut}^0(\text{Ab}(X))$, $\text{Aut}^0(F_2) = G_2^{lin}$ and $\text{Aut}^0(\text{Ab}(X)) = \text{Aut}^{Ab}(X) = G_1^{Ab} \times G_2^{Ab}$. Regrouping the factors, one has $X = (F_1 \times G_1^{Ab}) \times (F_2 \times G_2^{Ab}) = Y_1 \times Y_2$ and $\text{Aut}^0(Y_1) = G_1$, $\text{Aut}^0(Y_2) = G_2$. \square

Definition 6.2. We say that a homogeneous variety X is indecomposable if it is not the product of two varieties of dimension greater than zero.

Corollary 6.3. A parabolic variety X is indecomposable if and only if its automorphism group is a simple group of adjoint type.

Corollary 6.4. Any parabolic variety \mathcal{P} splits uniquely (up to permutation of the factors) as a direct product of indecomposable parabolic varieties, $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_r$.

Proof. This follows from Theorem 6.1 and from the existence and uniqueness of the decomposition of an affine semisimple group of adjoint type as a product of simple groups. \square

Remark 6.5. From Theorem 5.7 and Corollary 6.4, the classification of homogeneous varieties is reduced to the classification of abelian varieties and the classification of parabolic varieties with simple (of adjoint type) group of automorphisms.

Let G be an affine, simple, algebraic group of adjoint type. Let $\mathcal{B} = G/B$. Then $G = \text{Aut}^0(\mathcal{B})$ (see [4]), and any parabolic G -variety \mathcal{P} is an invariant quotient $\mathcal{B} \rightarrow \mathcal{P}$ (in a unique way up to translations by elements of G). The problem is to determine the parabolic G -varieties \mathcal{P} satisfying $\text{Aut}^0(\mathcal{P}) = G$.

First of all, let us specify what are the parabolic G -varieties, up to isomorphisms of varieties. Let us consider the associated parabolic varieties $\mathcal{P}_1 = G/P_1, \dots, \mathcal{P}_s = G/P_s$, where P_1, \dots, P_s are the maximal reduced parabolic subgroups containing B .

Given two parabolic G -varieties $\pi: \mathcal{B} \rightarrow \mathcal{P}$ and $\pi': \mathcal{B} \rightarrow \mathcal{P}'$, we shall denote by $\mathcal{P} * \mathcal{P}'$ the parabolic G -variety image of the G -morphism $\pi \times \pi': \mathcal{B} \rightarrow \mathcal{P} \times \mathcal{P}'$.

Remark 6.6. If $\mathcal{P} = G/P$ and $\mathcal{P}' = G/P'$ with $P, P' \subset G$ parabolic subgroups containing B , then $\mathcal{P} * \mathcal{P}' = G/P \cap P'$. Moreover, $\mathcal{P} * \mathcal{P}' = \sup(\mathcal{P}, \mathcal{P}')$, with respect to the order of the parabolic G -varieties.

Assume $\text{char}(k) = p > 0$. Given a variety X and a natural number $n \in \mathbb{N}$, we shall denote by $X^{[n]}$ the scheme whose underlying topological space is X and whose structural sheaf is the subsheaf $\mathcal{O}_X^{p^n} \subset \mathcal{O}_X$. One has a natural morphism of schemes $F^n: X \rightarrow X^{[n]}$ (F being the Frobenius morphism).

If G is a scheme in groups, then $G^{[n]}$ is a scheme in groups and $F^n: G \rightarrow G^{[n]}$ is a morphism of groups. We shall denote $G_n = \ker F^n$, which is a subscheme in groups, finite and local.

Remark 6.7. It is easy to see that if \mathcal{P} is a parabolic G -variety, then $\mathcal{P}^{[n]}$ is a parabolic G -variety too, and $\text{Aut}^0(\mathcal{P}^{[n]}) = \text{Aut}^0(\mathcal{P})^{[n]}$.

Theorem 6.8. Assume that $0 < \text{char}(k) \neq 2, 3$. For each parabolic G -variety \mathcal{P} , there exist unique indices $1 \leq i_1 < \dots < i_r \leq s$ and exponents $n_1, \dots, n_r \in \mathbb{N}$ such that

$$\mathcal{P} = \mathcal{P}_{i_1}^{[n_1]} * \dots * \mathcal{P}_{i_r}^{[n_r]}.$$

Moreover, $\text{Aut}^0(\mathcal{P}) = G$ if and only if $n_h = 0$ for some $1 \leq h \leq r$ and \mathcal{P}_{i_j} is non-exceptional for some $1 \leq j \leq r$. That is, $G \rightarrow \text{Aut}^0(\mathcal{P})$ is not an isomorphism if and only if either $n_1, \dots, n_r > 0$ or P is maximal and $P \subset G$ is an exceptional pair.

Proof. With the notation of [12], one has that each parabolic subgroup P of G containing B can be expressed in a unique way as $P = \bigcap_{\beta_i \in \mathcal{S}} P_{n_i, \beta_i}$, where \mathcal{S} is the basis of the root system of G corresponding to $T \subset B$, P_{β_i} is the maximal parabolic subgroup such that the root system of (P_{β_i}, T) does not contain $-\beta_i$, and $P_{n_i, \beta_i} = G_{n_i} \cdot P_{\beta_i}$ if $n_i \neq \infty$, and $P_{\infty, \beta_i} = G$. With our notation, $G/P_{\beta_i} = \mathcal{P}_i$, $G/P_{n_i, \beta_i} = \mathcal{P}_i^{[n_i]}$ if $n_i \neq \infty$, and $G/P_{\infty, \beta_i} = \text{Spec } k$; then $\mathcal{P} = G/P = \mathcal{P}_1^{[n_1]} * \dots * \mathcal{P}_s^{[n_s]}$, where $\mathcal{P}_i^{[\infty]} = \text{Spec } k$, i.e., it is a factor that can be suppressed.

For the second part, one observes that $G \rightarrow \text{Aut}^0(\mathcal{P})$ is injective if and only if $\inf(n_1, \dots, n_r) = 0$, since, on the contrary, it factors through $G^{[1]}$. Moreover, it is surjective if and only if there exists a morphism $\pi: \text{Aut}^0(\mathcal{P}) \rightarrow G^{[n]}$ that coincides with F^n over G . Indeed, if π exists, then $(\ker \pi)^0$ is a normal subgroup

of $\text{Aut}^0(\mathcal{P})$, and $(\ker \pi)^0 \cap G = \{e\}$. Hence, since $\text{Aut}^0(\mathcal{P})$ is simple, one concludes that $(\ker \pi)^0 = \{e\}$, and then $\dim \text{Aut}^0(\mathcal{P}) = \dim G^{[n]} = \dim G$; hence $G = \text{Aut}^0(\mathcal{P})$. Now, if (G, P_i) is non-exceptional, then $\text{Aut}^0(\mathcal{P}_i) = G$. So (by the following lemma) one has $\text{Aut}^0(\mathcal{P}) \rightarrow \text{Aut}^0(\mathcal{P}_i^{[n_i]}) = \text{Aut}^0(\mathcal{P}_i)^{[n_i]} = G^{[n_i]}$ and one concludes. Conversely, if P is maximal and the pair $P \subset G$ is exceptional, then $G \subsetneq \text{Aut}^0(\mathcal{P})$. \square

Lemma 6.9. *With the notation of the preceding theorem, the projections $\mathcal{P} \rightarrow \mathcal{P}_i^{[n_i]}$ are invariant quotients.*

Proof. Let us consider the G -morphism $\pi_i: \mathcal{P} \rightarrow \mathcal{P}_i^{[n_i]}$, let $\bar{\mathcal{L}}$ be a very ample invertible sheaf on $\mathcal{P}_i^{[n_i]}$ and $\mathcal{L} = \pi_i^* \bar{\mathcal{L}}$. Let $\bar{\pi}: \mathcal{P} \rightarrow \bar{\mathcal{P}}$ be the invariant quotient defined by the complete linear system $H^0(\mathcal{P}, \mathcal{L})$ (Proposition 4.3). It is clear that $\bar{\pi}$ factors through a morphism $h: \bar{\mathcal{P}} \rightarrow \mathcal{P}_i^{[n_i]}$ and $h^* \bar{\mathcal{L}}$ is a very ample invertible sheaf on $\bar{\mathcal{P}}$; hence $h: \bar{\mathcal{P}} \rightarrow \mathcal{P}_i^{[n_i]}$ is a finite morphism with connected fibers (since these fibers are quotients of parabolic subgroups of G). Then $\bar{\mathcal{P}} = \mathcal{P}_i^{[n_i]}$ and, by the preceding theorem (first part), $n \geq n_i$. It is clear that $\mathcal{P} = \mathcal{P}_1^{[n_1]} * \dots * \mathcal{P}_i^{[n_i]} * \dots * \mathcal{P}_s^{[n_s]}$; hence, again by the first part of the preceding theorem, $n = n_i$; that is, $\bar{\mathcal{P}} = \mathcal{P}_i^{[n_i]}$ and $\pi_i = h$ is an invariant quotient. \square

Let G be a simple algebraic group (of adjoint type) and \mathcal{D}_G its Dynkin diagram. It is a connected graph (the Coxeter graph, whose vertices are the elements of a basis of roots of G), where each edge has a weight $m_i = 1, 2, 3$. One has that $\text{Aut}(G)/\text{Inn}(G) = \text{Aut}(\mathcal{D}_G)$, i.e., the automorphisms of G , modulo inner automorphisms, are the automorphisms of the graph \mathcal{D}_G leaving the weights invariant. Moreover, one has that $\text{Aut } G = \text{Inn}(G) \rtimes \text{Aut}(\mathcal{D}_G)$, where the identification $\text{Aut}(\mathcal{D}_G) \subset \text{Aut } G$ is given by the automorphisms of G that leave invariant a given Borel subgroup B and a given maximal torus $T \subset B$. Therefore, $\text{Aut}(\mathcal{D}_G)$ acts on the set of parabolic subgroups containing B . The action is as follows: the vertices of \mathcal{D}_G correspond with the elements of a basis \mathcal{S} of roots of G , and these correspond with the reduced and maximal parabolic subgroups of G containing B , P_1, \dots, P_s . On the other hand, the parabolic subgroups of G containing B correspond biunivocally with the functions $\phi: \mathcal{S} \rightarrow \mathbb{N} \cup \infty$, i.e., with $\mathbb{N}^* \times \dots \times \mathbb{N}^* = \mathbb{N}^{*\mathcal{S}}$, where $\mathbb{N}^* = \mathbb{N} \cup \infty$ and $\text{Aut}(\mathcal{D}_G)$ acts on \mathcal{S} . Finally, for G fixed, the parabolic G -varieties, modulo isomorphisms of varieties, correspond with the parabolic subgroups of G , modulo automorphisms of G , i.e., with the parabolic subgroups of G containing B , modulo $\text{Aut}(\mathcal{D}_G)$.

In conclusion, in characteristic different from 2 and 3, one has:

Theorem 6.10 (Classification of indecomposable parabolic varieties). *Let G be a simple algebraic group (of adjoint type), \mathcal{S} the basis of roots of G corresponding to $T \subset B \subset G$, and \mathcal{D}_G its Dynkin diagram. The set of parabolic varieties whose automorphism group is isomorphic to G , modulo isomorphisms of varieties, is identified with the subset of the set*

$$\{\text{Parabolic } G\text{-Varieties}\} / \sim \quad \xrightarrow{\sim} \quad \mathbb{N}^{*\mathcal{S}} / \text{Aut}(\mathcal{D}_G)$$

formed by the classes of elements $(n_1, \dots, n_s) \in \mathbb{N}^{\mathcal{S}}$ such that $n_i = 0$ for some i and $n_j \neq \infty$ for some j such that $P_j \subset G$ is non-exceptional.*

ACKNOWLEDGMENTS

The author thanks the referee for valuable comments and suggestions.

REFERENCES

1. Borel, A. *Symmetric Compact Complex Spaces*. Arch. Math. (Basel) **33** (1979/80), no. 1, 49–56. MR **80k**:32033
2. Borel, A. and Remmert, R. *Über kompakte homogene Kählersche Mannigfaltigkeiten*. Math. Ann. **145** (1961/1962), no. 1, 429–439. MR **26**:3088
3. Chevalley, C. *Séminaire sur la Classification des Groupes de Lie Algébriques*. Paris: Ecole Norm. Sup. 1956–1958. MR **21**:5696
4. Demazure, M. *Automorphismes et Déformations des Variétés de Borel*. Invent. Math. **39**, 179–186 (1977). MR **55**:8054
5. Grothendieck, A. *Technique de descente et théorèmes d'existence en géométrie algébrique. V: Les schémas de Picard: Théorèmes d'existence*, Séminaire Bourbaki, 14ième année, 1961/62, fasc. 3, Exposé 232, Secrétariat Math., Paris, 1962, and reprints. MR **26**:3561; MR **33**:5420i; MR **99f**:00039
6. Humphreys, J. E. *Linear Algebraic Groups*. Graduate Texts in Mathematics **21**, Springer-Verlag, New York (1975). MR **53**:633
7. Matsumura, H. and Oort, F. *Representability of Group Functors, and Automorphisms of Algebraic Schemes*. Invent. Math. **4**, 1–25 (1967). MR **36**:181
8. Mumford, D. *Abelian Varieties*. Tata Studies in Math., Oxford Univ. Press (1970). MR **44**:219
9. Mumford, D. *On the Equations Defining Abelian Varieties, I, II, III*. Invent. Math. **1** (1966), 287–354; **3** (1967), 76–135, 215–244. MR **34**:4269; MR **36**:2621; MR **36**:2622
10. Mumford, D. and Fogarty, J. *Geometric Invariant Theory*. Springer-Verlag (1982). MR **86a**:14006
11. Rosenlicht, M. *Some Basic Theorems on Algebraic Groups*. Amer. J. of Math. **78**, 427–443, (1956). MR **18**:514a
12. Wenzel, Ch. *Classification of all Parabolic Subgroup-Schemes of a Reductive Linear Algebraic Group over an Algebraically Closed Field*. Trans. Amer. Math. Soc. **337**, 211–218 (1993). MR **93g**:20090

DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD DE SALAMANCA, PLAZA DE LA MERCED 3-4,
C.P. 37008, ESPAÑA

E-mail address: `sancho@gugu.usal.es`

A PATH-TRANSFORMATION FOR RANDOM WALKS AND THE ROBINSON-SCHENSTED CORRESPONDENCE

NEIL O'CONNELL

ABSTRACT. The author and Marc Yor recently introduced a path-transformation $G^{(k)}$ with the property that, for X belonging to a certain class of random walks on \mathbb{Z}_+^k , the transformed walk $G^{(k)}(X)$ has the same law as the original walk conditioned never to exit the Weyl chamber $\{x : x_1 \leq \cdots \leq x_k\}$. In this paper, we show that $G^{(k)}$ is closely related to the Robinson-Schensted algorithm, and use this connection to give a new proof of the above representation theorem. The new proof is valid for a larger class of random walks and yields additional information about the joint law of X and $G^{(k)}(X)$. The corresponding results for the Brownian model are recovered by Donsker's theorem. These are connected with Hermitian Brownian motion and the Gaussian Unitary Ensemble of random matrix theory. The connection we make between the path-transformation $G^{(k)}$ and the Robinson-Schensted algorithm also provides a new formula and interpretation for the latter. This can be used to study properties of the Robinson-Schensted algorithm and, moreover, extends easily to a continuous setting.

1. INTRODUCTION AND SUMMARY

For $k \geq 2$, denote the set of probability distributions on $\{1, \dots, k\}$ by \mathcal{P}_k . Let $(\xi_m, m \geq 1)$ be a sequence of independent random variables with common distribution $p \in \mathcal{P}_k$ and, for $1 \leq i \leq k, n \geq 0$, set

$$(1) \quad X_i(n) = |\{1 \leq m \leq n : \xi_m = i\}|.$$

If $p_1 < \cdots < p_k$, there is a positive probability that the random walk $X = (X_1, \dots, X_k)$ never exits the Weyl chamber

$$(2) \quad W = \{x \in \mathbb{R}^k : x_1 \leq \cdots \leq x_k\};$$

this is easily verified using, for example, the concentration inequality

$$P(|X(n) - np| > \varepsilon) \leq K e^{-c(\varepsilon)n},$$

where $c(\varepsilon)$ and K are finite positive constants.

Received by the editors March 7, 2002 and, in revised form, October 25, 2002.

2000 *Mathematics Subject Classification.* Primary 05E05, 05E10, 15A52, 60B99, 60G50, 60J27, 60J45, 60J65, 60K25, 82C41.

Key words and phrases. Pitman's representation theorem, random walk, Brownian motion, Weyl chamber, Young tableau, Robinson-Schensted correspondence, RSK, intertwining, Markov functions, Hermitian Brownian motion, random matrices.

In [36], a certain path-transformation $G^{(k)}$ was introduced with the property that:

Theorem 1.1. *The law of the transformed walk $G^{(k)}(X)$ is the same, assuming $p_1 < \dots < p_k$, as that of the original walk X conditioned never to exit W .*

We will recall the definition of $G^{(k)}$ in Section 2 below.

This was motivated by a desire to find a multi-dimensional generalisation of Pitman's representation for the three-dimensional Bessel process [37], and to understand some striking connections which were recently discovered by Baik, Deift and Johansson [4], Baryshnikov [5] and Gravner, Tracy and Widom [21], between oriented percolation and random matrices. For more background on this, see [33].

The proof of Theorem 1.1 given in [36] uses certain symmetry and reversibility properties of M/M/1 queues in series; consequently, the transformation $G^{(k)}$ has a "queueing-theoretic" interpretation.

In this paper we will show that the path-transformation $G^{(k)}$ is closely related to the Robinson-Schensted correspondence. More precisely, if $\lambda(n) = (\lambda_1(n) \geq \dots \geq \lambda_k(n))$ denotes the shape of the Young tableaux obtained, when one applies the Robinson-Schensted algorithm with column-insertion, to the random word $\xi_1 \dots \xi_n$, then (for any realisation of X)

$$(G^{(k)}(X))(n) = (\lambda_k(n), \dots, \lambda_1(n)).$$

Immediately, this yields a new representation and formula for the Robinson-Schensted algorithm, and this formula has a queueing interpretation. We will use this representation to recover known, and perhaps not-so-well-known, properties of the Robinson-Schensted algorithm.

Given this connection, Theorem 1.1 can now be interpreted as a statement about the evolution of the shape $\lambda(n)$ of a certain randomly growing Young tableau. We give a direct proof of this result using properties of the Robinson-Schensted correspondence. This also yields more information about the joint law of X and $G^{(k)}(X)$, and dispenses with the condition $p_1 < \dots < p_k$.

As in [36], the corresponding results for the Brownian motion model can be recovered by Donsker's theorem. The path-transformation $G^{(k)}$ extends naturally to a continuous setting and, given the connection with the Robinson-Schensted algorithm, the continuous version can now be regarded as a natural extension of the Robinson-Schensted algorithm to a continuous setting. As discussed in [36], the results for Brownian motion have an interpretation in random matrix theory. In particular, Theorem 1.1 yields a representation for the eigenvalue process associated with Hermitian Brownian motion as a certain path-transformation (the continuous analogue of $G^{(k)}$) applied to a standard Brownian motion. The new results presented in this paper also yield new results in this context. This random matrix connection comes from the well-known fact that the eigenvalue process associated with Hermitian Brownian motion can be interpreted as Brownian motion conditioned never to exit the Weyl chamber W . We remark that a similar representation for the eigenvalues of Hermitian Brownian motion was independently obtained by Bougerol and Jeulin [11], in a more general context, by completely different methods.

The outline of the paper is as follows. In the next section we recall the definition of $G^{(k)}$ and record some of its properties. In section 3, we make the connection with

the Robinson-Schensted algorithm, and briefly consider some immediate implications of this connection. A worked example is presented in section 4. In section 5, we record some properties of the conditioned walk of Theorem 1.1 and extend its definition beyond the case $p_1 < \dots < p_k$. In section 6, we prove a generalisation of Theorem 1.1, in the context of Young tableaux, using properties of the RS correspondence. In section 7, we define a continuous version of the path-transformation and present the ‘‘Poissonized’’ analogues of the results of the previous section. In section 8, we present the corresponding results for the Brownian model, and briefly discuss the connection with random matrices. An application in queueing theory is presented in section 9, and we conclude the paper with some remarks in section 10.

Some notation. Let $b = \{e_1, \dots, e_k\}$ denote the standard basis elements in \mathbb{R}^k . For $x, y \in \mathbb{R}_+^k$ we will write $x^y = x_1^{y_1} \dots x_k^{y_k}$, $xy = (x_1 y_1, \dots, x_k y_k)$, $|x| = \sum_i x_i$ and define $x^* \in \mathbb{R}_+^k$ by $x_i^* = x_{k-i+1}$. Denote the origin in \mathbb{R}^k by o .

Acknowledgements. Thanks to Francois Baccelli, Phillipe Biane, Phillipe Bougerol and Marc Yor for many helpful and illuminating discussions on these topics. This research was partly carried out during a visit, funded by the CNRS, to the Laboratoire de Probabilités, Université Paris 6, and partly at the ENS, thanks to financial support of INRIA.

2. THE PATH-TRANSFORMATION

The support of the random walk X , which we denote by Π_k , consists of paths $x : \mathbb{Z}_+ \rightarrow \mathbb{Z}_+^k$ with $x(0) = 0$ and, for each $n > 0$, $x(n) - x(n-1) \in b$. Let Π_k^W denote the subset of those paths taking values in W . It is convenient to introduce another set Λ_k of paths $x : \mathbb{Z}_+ \rightarrow \mathbb{Z}_+^k$ with $x(0) = 0$ and $x(n) - x(n-1) \in \{0, e_1, \dots, e_k\}$, for each $n > 0$.

For $x, y \in \Lambda_1$, define $x \triangle y \in \Lambda_1$ and $x \nabla y \in \Lambda_1$ by

$$(3) \quad (x \triangle y)(n) = \min_{0 \leq m \leq n} [x(m) + y(n) - y(m)]$$

and

$$(4) \quad (x \nabla y)(n) = \max_{0 \leq m \leq n} [x(m) + y(n) - y(m)].$$

The operations \triangle and ∇ are not associative in general. Unless otherwise delineated by parentheses, the default order of operations is from left to right; for example, when we write $x \triangle y \triangle z$, we mean $(x \triangle y) \triangle z$.

The mappings $G^{(k)} : \Lambda_k \rightarrow \Lambda_k$ are defined as follows. Set

$$(5) \quad G^{(2)}(x, y) = (x \triangle y, y \nabla x)$$

and, for $k > 2$,

$$(6) \quad G^{(k)}(x_1, \dots, x_k) = (x_1 \triangle x_2 \triangle \dots \triangle x_k, \\ G^{(k-1)}(x_2 \nabla x_1, x_3 \nabla (x_1 \triangle x_2), \dots, x_k \nabla (x_1 \triangle \dots \triangle x_{k-1}))).$$

Note that $G^{(k)} : \Pi_k \rightarrow \Pi_k^W$.

We will now give an alternative definition of $G^{(k)}$ which will be useful for making the connection with the Robinson-Schensted correspondence.

Occasionally, we will suppress the dependence of functions on x , when the context is clear: for example, we may write $G^{(k)}$ instead of $G^{(k)}(x)$, and so on.

For $k \geq 2$, define maps $D^{(k)} : \Lambda_k \rightarrow \Lambda_k$ and $T^{(k)} : \Lambda_k \rightarrow \Lambda_{k-1}$ by

$$(7) \quad D^{(k)}(x) = (x_1, x_1 \triangle x_2, \dots, x_1 \triangle \dots \triangle x_k)$$

and

$$(8) \quad T^{(k)}(x) = (x_2 \nabla x_1, x_3 \nabla (x_1 \triangle x_2), \dots, x_k \nabla (x_1 \triangle \dots \triangle x_{k-1})).$$

For notational convenience, let $D^{(1)}$ be the identity transformation.

Note that the above definition is recursive: for $i \geq 2$,

$$(9) \quad D_i^{(k)} = D_{i-1}^{(k)} \triangle x_i$$

and

$$(10) \quad T_{i-1}^{(k)} = x_i \nabla D_{i-1}^{(k)}.$$

Alternatively, we can write

$$(11) \quad (D_i^{(k)}, T_{i-1}^{(k)}) = G^{(2)}(D_{i-1}^{(k)}, x_i).$$

For each $x \in \Lambda_k$, consider the triangular array of sequences $d^{(i)} \in \Lambda_{k-i+1}$, $1 \leq i \leq k$, defined as follows. Set

$$\begin{aligned} d^{(1)} &= D^{(k)}(x), & t^{(1)} &= T^{(k)}(x), \\ d^{(2)} &= D^{(k-1)}(t^{(1)}), & t^{(2)} &= T^{(k-1)}(t^{(1)}), \end{aligned}$$

and so on; for $i \leq k$,

$$d^{(i)} = D^{(k-i+1)}(t^{(i-1)}),$$

and for $i \leq k-1$,

$$t^{(i)} = T^{(k-i+1)}(t^{(i-1)}).$$

Recalling the definition of $G^{(k)}$ given earlier, we see that

$$(12) \quad G^{(k)} = (d_k^{(1)}, \dots, d_1^{(k)}).$$

Note also that, for each $i \leq k$,

$$(13) \quad G^{(i)}(x_1, \dots, x_i) = (d_i^{(1)}, \dots, d_1^{(i)}).$$

We will conclude this section by recording some useful properties and interpretations of the operations \triangle and ∇ , and of the path-transformation $G^{(k)}$, for later reference. We defer the proofs: these will be given in the appendix.

The following notation for increments of paths will be useful: for $x \in \Lambda_k$ and $l \geq n$, set $x(n, l) = x(l) - x(n)$.

The operations \triangle and ∇ have a queueing-theoretic interpretation, which we will make strong use of when we make the connection with the Robinson-Schensted correspondence in the next section. For more general discussions on "min-plus algebra" and queueing networks, see [1].

Suppose $(x, y) \in \Pi_2$, and consider a simple queue which evolves as follows. At each time n , either $x(n) - x(n-1) = 1$ and $y(n) - y(n-1) = 0$, in which case a new customer arrives at the queue, or $x(n) - x(n-1) = 0$ and $y(n) - y(n-1) = 1$, in which case, if the queue is not empty, a customer departs (otherwise nothing happens). The number of customers remaining in the queue at time n , which we denote by $q(n)$, satisfies the Lindley recursion

$$(14) \quad q(n) = \max\{q(n-1) + \epsilon(n), 0\},$$

where $\epsilon(n) = x(n) - x(n-1) - y(n) + y(n-1)$. Iterating (14), we obtain

$$(15) \quad q(n) = \max_{0 \leq m \leq n} [x(m, n) - y(m, n)].$$

Thus, the number of customers $d(n)$ to depart up to and including time n is given by

$$(16) \quad d(n) = x(n) - q(n) = (x \triangle y)(n).$$

We also have

$$(17) \quad t(n) := x(n) + u(n) = (y \nabla x)(n),$$

where

$$u(n) = y(n) - d(n)$$

is the number of times $m \leq n$ that $y(m) - y(m-1) = 1$ and $q(m-1) = 0$; in the language of queueing theory, $u(n)$ is the number of “unused services” up to and including time n . (For this queue we refer to the points of increase of y as “services”.)

Lemma 2.1. For $(x, y) \in \Lambda_2$,

$$(18) \quad x \triangle y + y \nabla x = x + y$$

and, if $\min_{l \geq n} [x(l) - (x \triangle y)(l)] = 0$,

$$\begin{aligned} x(n) - (x \triangle y)(n) &= \max_{0 \leq m \leq n} [x(m, n) - y(m, n)] \\ &= \max_{l \geq n} [(x \triangle y)(n, l) - (y \nabla x)(n, l)]. \end{aligned}$$

In particular, writing $G^{(2)} \equiv G^{(2)}(x, y)$, we have

$$(19) \quad (x(n), y(n)) = G^{(2)}(n) + F^{(2)} \left(G^{(2)}(n, l), l \geq n \right),$$

where $F^{(2)} : \mathcal{D} \rightarrow \mathbb{Z}^2$ is defined on

$$\mathcal{D} = \{z \in (\mathbb{Z}^2)^{\mathbb{Z}_+} : M(z) = \max_{n \geq 0} [z_1(n) - z_2(n)] < \infty\}$$

by $F^{(2)}(z) = (M(z), -M(z))$.

In the queueing context described above, Lemma 2.1 states that $x + y = d + t$ and, if $\min_{l \geq n} q(l) = 0$,

$$(20) \quad q(n) = \max_{m \geq n} [d(n, m) - t(n, m)].$$

The first identity is readily verified. The formula for $q(n)$ in terms of the future increments of d and t follows from the time-reversal symmetry in the dynamics of the system: this formula is the dual of (15). When time is reversed, the roles played by (x, y) and (d, t) are interchanged. This symmetry is at the heart of the proof of Theorem 1.1 given in [36], where it is considered in an equilibrium context.

Note that, if we set $z = y - x$ and $s(n) = \max_{0 \leq m \leq n} z(m)$, then

$$y \nabla x - x \triangle y = 2s - z$$

and (20) is equivalent to the well-known identity

$$s(n) = \min_{l \geq n} [2s(l) - z(l)].$$

This is familiar in the context of Pitman’s representation for the three-dimensional Bessel process. Observe that the statement of Theorem 1.1 in the case $k = 2$ is

equivalent to the following discrete version of Pitman's theorem: if $\{Z(n), n \geq 0\}$ is a simple random walk on \mathbb{Z} with positive drift, started at 0, and we set $S(n) = \max_{0 \leq m \leq n} Z(m)$, then $2S - Z$ has the same law as that of Z conditioned to stay nonnegative. The usual statement of Pitman's theorem can be recovered from Theorem 8.1 below.

Lemma 2.1 has the following generalisation:

Lemma 2.2. *For $x \in \Lambda_k$, writing $G^{(k)} \equiv G^{(k)}(x)$, we have*

$$(21) \quad |G^{(k)}| = |x|$$

and, if $\min_{l \geq n} (d_j^{(i)} - d_{j+1}^{(i)})(l) = 0$, for $1 \leq j < i < k$ (c_n),

$$(22) \quad x(n) = G^{(k)}(n) + F^{(k)}\left(G^{(k)}(n, l), l \geq n\right),$$

where the function $F^{(k)}$ will be defined in the proof.

As we remarked earlier, the operations \triangle and ∇ are not associative. The following identities are useful for manipulating complex combinations of these operations.

Lemma 2.3. *For $(a, b, c) \in \Lambda_3$ we have*

$$(23) \quad a \nabla (c \triangle b) \nabla (b \nabla c) = a \nabla b \nabla c$$

and

$$(24) \quad a \triangle (c \nabla b) \triangle (b \triangle c) = a \triangle b \triangle c.$$

For example, (23) immediately yields

Lemma 2.4. *For $x \in \Lambda_k$, $G_k^{(k)}(x) = x_k \nabla \cdots \nabla x_1$.*

3. CONNECTION WITH THE ROBINSON-SCHENSTED ALGORITHM

We refer the reader to the books of Fulton [18] and Stanley [40] for detailed discussions on the Robinson-Schensted algorithm and its properties. The standard Robinson-Schensted algorithm takes a word $w = a_1 \cdots a_n \in \{1, 2, \dots, k\}^n$ and proceeds, by "row-inserting" the numbers a_1 , then a_2 , and so on, to construct a semistandard tableau $P(w)$ associated with w , of size n with entries from the set $\{1, 2, \dots, k\}$. If one also maintains a "recording tableau" $Q(w)$, which is a standard tableau of size n , the mapping from words $\{1, 2, \dots, k\}^n$ to pairs of semistandard and standard tableaux of size n , the semistandard tableau having entries from $\{1, 2, \dots, k\}$ and both having the same shape, is a bijection: this is the Robinson-Schensted correspondence. One can also do all of the above using "column-insertion" instead of row-insertion to construct the semistandard tableau, but still maintaining a recording tableau, and the resulting map is also a bijection. Column and row insertion are not the same thing, but they are related in the following way: the semistandard tableau obtained by applying the Robinson-Schensted algorithm, with column-insertion, to the word $a_1 \cdots a_n$ is the same as the one obtained by applying the Robinson-Schensted algorithm, with row-insertion, to the reversed word $a_n \cdots a_1$. The standard tableaux obtained in each case are also related, but we do not need this and refer the reader to [18] for details.

Fix $x \in \Pi_k$, let $d^{(i)} \in \Lambda_{k-i+1}$, $1 \leq i \leq k$, be the corresponding triangular array of sequences defined in the previous section.

For each n , construct a semistandard Young tableau as follows. In the first row, put

$$d_1^{(1)}(n) \text{ 1's, } d_1^{(2)}(n) - d_1^{(1)}(n) \text{ 2's, } \dots, d_1^{(k)}(n) - d_1^{(k-1)}(n) \text{ } k\text{'s;}$$

in the second row, put

$$d_2^{(1)}(n) \text{ 2's, } d_2^{(2)}(n) - d_2^{(1)}(n) \text{ 3's, } \dots, d_2^{(k-1)}(n) - d_2^{(k-2)}(n) \text{ } k\text{'s,}$$

and so on. In the final row, there are just $d_k^{(1)}(n)$ k 's. Denote this tableau by $\tau(n)$. For example, if $k = 3$ and

$$(25) \quad \begin{array}{ccccc} d_1^{(1)}(7) & d_2^{(1)}(7) & d_3^{(1)}(7) & 2 & 2 & 1 \\ & d_1^{(2)}(7) & d_2^{(2)}(7) & = & 3 & 2 \\ & & d_1^{(3)}(7) & & & 4 \end{array}$$

then the corresponding semistandard tableau $\tau(7)$ is

$$(26) \quad \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 3 \\ \hline 2 & 2 & & \\ \hline 3 & & & \\ \hline \end{array}$$

Let a_m be the sequence defined by $a_m = i$ whenever

$$x(m) - x(m-1) = e_i.$$

Theorem 3.1. *The semistandard tableau $\tau(n)$ is precisely the one that is obtained when one applies the Robinson-Schensted algorithm, with column insertion, to the word $a_1 \cdots a_n$. In particular, if $l(n)$ denotes the shape of $\tau(n)$, then $l(n)^* = (G^{(k)}(x))(n)$.*

Proof. It will suffice to describe how the mapping $G^{(k)}$ acts on a typical element of Π_k , from an algorithmic point of view.

For each $k \geq 2$, the maps $D^{(k)} : \Lambda_k \rightarrow \Lambda_k$ and $T^{(k)} : \Lambda_k \rightarrow \Lambda_{k-1}$ can be defined as follows. Fix $x \in \Pi_k$ and set $d = D^{(k)}(x)$, $t = T^{(k)}(x)$. Set $d(0) = t(0) = 0$, and define the sequences $d(n)$ and $t(n)$ inductively on n . Suppose $x(n) - x(n-1) = e_i$; that is, $a_n = i$. We need to treat the cases $i = 1$ and $i = k$ separately.

Suppose $i = 1$. Then we set $d(n) = d(n-1) + e_1$ and $t(n) = t(n-1) + e_1$.

If $i = k$, and $d_k(n-1) < d_{k-1}(n-1)$, we set $d(n) = d(n-1) + e_k$ and $t(n) = t(n-1)$.

If $i = k$, and $d_k(n-1) = d_{k-1}(n-1)$, we set $d(n) = d(n-1)$ and $t(n) = t(n-1) + e_{k-1}$.

Now suppose $1 < i < k$. If $d_i(n-1) < d_{i-1}(n-1)$, set $d(n) = d(n-1) + e_i$ and $t(n) = t(n-1) + e_i$; if $d_i(n-1) = d_{i-1}(n-1)$, set $d(n) = d(n-1)$ and $t(n) = t(n-1) + e_{i-1}$. Recall that $D^{(1)}$ is the identity transformation.

In queueing language, we have just constructed a series of k queues in tandem. Initially there are infinitely many customers in the first queue and the other queues are all empty. At each time n , if $x_i(n+1) - x_i(n) = 1$ (or, equivalently, $a_n = i$) there is a "service event" at the i^{th} queue; if this queue is not empty a customer departs from it and, if $i < k$, joins the $(i+1)^{\text{th}}$ queue. The number of departures from the i^{th} queue up to and including time n is given by $d_i(n)$ and $t_i(n) = d_i(n) + u_i(n)$, where $u_i(n)$ is the number of "unused services" at the $(i+1)^{\text{th}}$ queue up to and including time n . Recalling the queueing-theoretic interpretations of Δ and ∇ , we see that $d_i = x_1 \Delta \cdots \Delta x_i$ and $t_i = x_i \nabla d_{i-1}$.

Now fix $x \in \Pi_k$, and recall the definition of the triangular array of sequences $d^{(i)} \in \Lambda_{k-i+1}$, $1 \leq i \leq k$. Set

$$\begin{aligned} d^{(1)} &= D^{(k)}(x), & t^{(1)} &= T^{(k)}(x), \\ d^{(2)} &= D^{(k-1)}(t^{(1)}), & t^{(2)} &= T^{(k-1)}(t^{(1)}), \end{aligned}$$

and so on; for $i \leq k$,

$$d^{(i)} = D^{(k-i+1)}(t^{(i-1)}),$$

and for $i \leq k - 1$,

$$t^{(i)} = T^{(k-i+1)}(t^{(i-1)}).$$

Here we have constructed a “series of queues in series”, the entire system “driven” by x . This is represented in Figure 1 for the case $k = 3$.

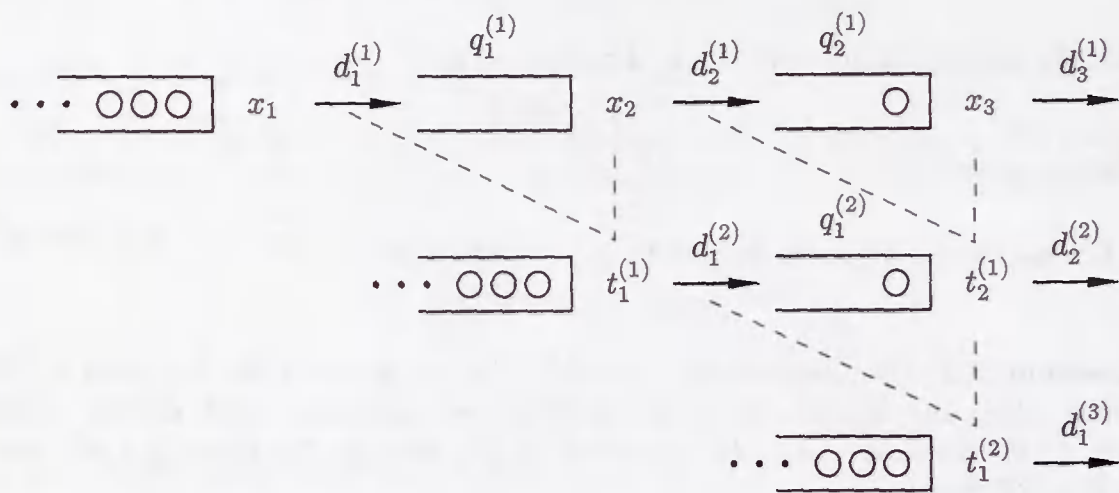


FIGURE 1. The series of queues in series ($k = 3$)

The first series of queues is just the one described above: there are k queues in the series, and initially queues 2 thru k are empty and the first queue has infinitely many customers; whenever x_i increases by one there is a service at the i^{th} queue and one customer is permitted to depart (and proceed to the next queue if $i < k$). The number of departures from the i^{th} queue up to time n is given by $d_i^{(1)}(n)$.

The second series of queues has $t^{(1)}$ “moving” the customers in place of x . This time there are $k - 1$ queues. Initially, queues 2 thru $k - 1$ are empty and the first queue has infinitely many customers. There is a service event at the i^{th} queue whenever $t_i^{(1)}$ increases by one—that is, whenever, in the first series, there is a departure from the i^{th} queue or an unused service at the $(i + 1)^{th}$ queue (these events will never occur simultaneously).

The second series generates a new sequence of t ’s, which we denote by $t^{(2)}$, and this is used to drive the third series, which consists of $k - 2$ queues, and so on.

It is useful to define

$$(27) \qquad q_i^{(j)} = d_i^{(j)} - d_{i+1}^{(j)};$$

$q_i^{(j)}(n)$ is just the number of customers in the $(i + 1)^{th}$ queue of the j^{th} series at time n . For example, in the network shown in Figure 1, $q_1^{(1)} = 0$ and $q_2^{(1)} = q_1^{(2)} = 1$.

Now consider the evolution of the corresponding semistandard tableaux $\tau(n)$, $n \geq 1$. See Figure 2. Another look at the algorithm described above should

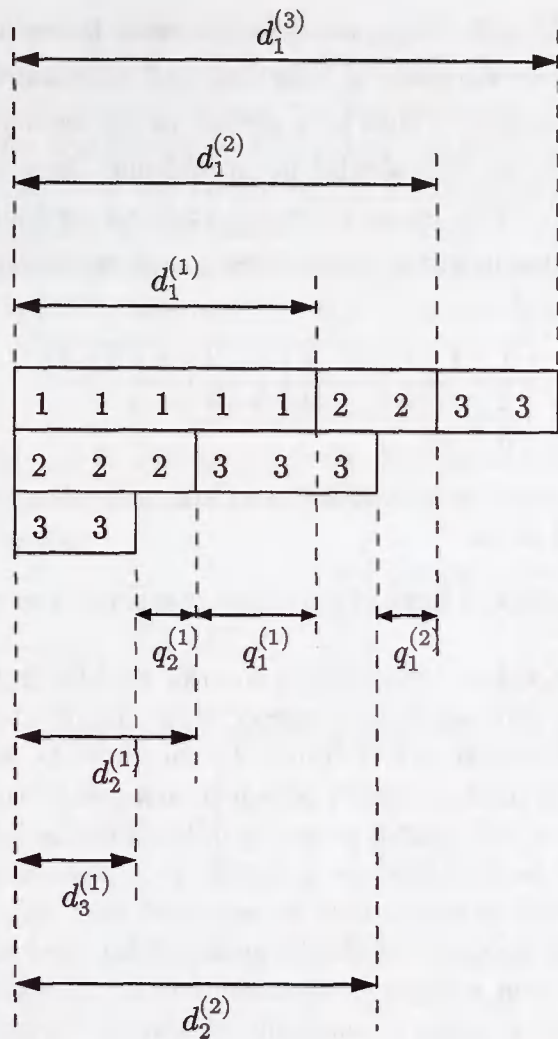


FIGURE 2. The tableau $\tau(17)$

convince the reader that $\tau(n)$ is precisely the semistandard tableau obtained when one applies the Robinson-Schensted algorithm, with *column* insertion, to the word $a_1 \cdots a_n$.

To see this, look at the tableau $\tau(17)$ represented in Figure 2. Recall that $a_m = i$ if x_i increases by one at time m , in which case there is a service at the i^{th} queue of the first series.

Suppose that the next “letter” $a(18) = 2$. Since $d_2^{(1)} < d_1^{(1)}$, that is, $q_1^{(1)} > 0$, we have a departure from the second queue in the first series; that is, we decrease $q_1^{(1)}$ by one and increase $d_2^{(1)}$ by one. In turn, this leads to an increase in $t_2^{(1)}$, that is, a service at the second queue in the second series, and so, since $q_1^{(2)} > 0$, we decrease $q_1^{(2)}$ by one, and increase $d_2^{(2)}$ by one. That’s it; the resulting tableau $\tau(18)$ is

1	1	1	1	1	2	2	3	3
2	2	2	2	3	3	3		
3	3							

and we recognise this procedure as column-insertion of the number 2 into the tableau $\tau(17)$.

Note that, now, $q_1^{(2)} = 0$. Suppose that the next letter $a(19)$ is also a 2. We still have $q_1^{(1)} > 0$; so we decrease $q_1^{(1)}$ by one and increase $d_2^{(1)}$ by one. In turn, this leads to an increase in $t_2^{(1)}$, that is, a service at the second queue in the second series; but $q_1^{(2)} = 0$, and so this service is unused and there is no departure (that is, no increase in $d_2^{(2)}$). The unused service leads to an increase in $t_1^{(3)}$, that is, a service at the only queue in the third series, and we increase $d_1^{(3)}$ by one. The resulting tableau $\tau(19)$ is

1	1	1	1	1	2	2	3	3	3
2	2	2	2	2	3	3			
3	3								

and, again, we recognise this procedure as column-insertion of the number 2 into the tableau $\tau(18)$, and so on. \square

We will give a completely worked example, starting from an empty tableau, in the next section.

We conclude this section with some remarks on the immediate implications of Theorem 3.1. Let $l(n)$ and $\alpha(n)$ respectively denote the shape and weight of $\tau(n)$. In view of Theorem 3.1, Lemma 2.4 can now be interpreted as stating that $l_1(n)$ is the length of the longest non-decreasing subsequence in the reversed word $a_n \cdots a_1$. This is a well-known property of the Robinson-Schensted algorithm. Thus, Lemma 2.4 can be regarded as a corollary of Theorem 3.1, or the proof of Lemma 2.4 given in the appendix can be regarded as a new proof of the longest increasing subsequence property of the Robinson-Schensted algorithm.

More generally, we can compare the statement of Theorem 3.1 with Greene's theorem, and this leads to some remarkable identities. Greene's theorem (see, for example, [29]) states that, if $m_i(n)$ denotes the maximum of the sum of the lengths of i disjoint, non-decreasing subsequences in the reversed word $a_n \cdots a_1$, then, for $i \leq k$,

$$(28) \quad m_i(n) = l_1(n) + l_2(n) + \cdots + l_i(n).$$

It therefore follows from Theorem 3.1 that

$$(29) \quad m_i(n) = G_k^{(k)}(n) + G_{k-1}^{(k)}(n) + \cdots + G_{k-i+1}^{(k)}(n).$$

It would be interesting to see a direct proof of this identity. Similarly, one can compare Theorem 3.1 with the various extensions of Greene's theorem given, for example, in [6].

The implications of Lemma 2.2 for the Robinson-Schensted algorithm would appear to be less well known. Fix $k \geq 2$, and set $H(z) = F^{(k)}(z^*)^*$.

Corollary 3.2. *Suppose that (c_n) holds. The weight $\alpha(n)$ of $\tau(n)$ can be recovered from the sequence of shapes*

$$\{\text{sh } \tau(l), \ l \geq n\}.$$

In fact, if we set, for $m \geq 0$,

$$u(m) = \text{sh } \tau(n+m) - \text{sh } \tau(n),$$

then

$$\alpha(n) = \text{sh } \tau(n) + H(u).$$

Recall that the recording tableaux $\sigma(n)$ are nested; the limiting standard tableau $\sigma(\infty) = \lim_{n \rightarrow \infty} \sigma(n)$ is thus a well-defined object. It follows from Corollary 3.2 that provided (c_n) holds we can recover the infinite word $a_1 a_2 \dots$ from $\sigma(\infty)$.

This is also true for the Robinson-Schensted algorithm applied with row insertion. To see this, recall that the recording tableaux maintained when one applies the row insertion algorithm to the infinite word $a_1 a_2 \dots$ is the same as that maintained when one applies the column insertion algorithm to the word $a_1^\dagger a_2^\dagger \dots$, where $a_n^\dagger = k - a_n + 1$ (or see, for example, [18, A.2, Exercise 5]).

4. A WORKED EXAMPLE

Suppose $k = 3$ and $n = 7$, and we apply the Robinson-Schensted algorithm with column insertion to the word $a_1 \cdots a_7 = 3112322$. We obtain the following sequence of semistandard tableaux:

<table><tr><td>3</td></tr></table>	3	<table><tr><td>1</td><td>3</td></tr></table>	1	3	<table><tr><td>1</td><td>1</td><td>3</td></tr></table>	1	1	3	<table><tr><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td></td><td></td></tr></table>	1	1	3	2			<table><tr><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td></td><td></td></tr><tr><td>3</td><td></td><td></td></tr></table>	1	1	3	2			3			<table><tr><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td>2</td><td></td></tr><tr><td>3</td><td></td><td></td></tr></table>	1	1	3	2	2		3			<table><tr><td>1</td><td>1</td><td>2</td><td>3</td></tr><tr><td>2</td><td>2</td><td></td><td></td></tr><tr><td>3</td><td></td><td></td><td></td></tr></table>	1	1	2	3	2	2			3			
3																																																
1	3																																															
1	1	3																																														
1	1	3																																														
2																																																
1	1	3																																														
2																																																
3																																																
1	1	3																																														
2	2																																															
3																																																
1	1	2	3																																													
2	2																																															
3																																																

The evolution of the corresponding queueing network is as follows. For each n , set

$$Q(n) = \begin{bmatrix} q_1^{(1)}(n) & q_2^{(1)}(n) \\ & q_1^{(2)}(n) \end{bmatrix}$$

and

$$D(n) = \begin{bmatrix} d_1^{(1)}(n) & d_2^{(1)}(n) & d_3^{(1)}(n) \\ & d_1^{(2)}(n) & d_2^{(2)}(n) \\ & & d_1^{(3)}(n) \end{bmatrix}.$$

Initially,

$$Q(0) = \begin{bmatrix} 0 & 0 \\ & 0 \end{bmatrix} \quad \text{and} \quad D(0) = \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix}.$$

At time 1, $a_1 = 3$; so there is a service at the third queue in the first series. Since $q_2^{(1)}(0) = 0$, this queue is empty, and the service is unused, leading to an increase in $t_2^{(1)}$ and hence a service at the second queue in the second series. This is also unused; so we have a service at the first (and only) queue in the third series, and there is a departure there: $d_1^{(3)}(1) = 1$. Thus,

$$Q(1) = \begin{bmatrix} 0 & 0 \\ & 0 \end{bmatrix} \quad \text{and} \quad D(1) = \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 1 \end{bmatrix}.$$

The corresponding tableau $\tau(1)$ is

<table border="1"><tr><td>3</td></tr></table>	3
3	

At time 2, $a_2 = 1$; so there is a service at the first queue in the first series and a customer departs to join the second queue: thus, $d_1^{(1)}(2) = 1$ and $q_1^{(1)}(2) = 1$. In the second series, this leads to an increase in $t_1^{(1)}$ and hence a departure from the first queue to the second queue: thus, $d_1^{(2)}(2) = 1$ and $q_1^{(2)}(2) = 1$. In turn, this

leads to an increase in $t_1^{(2)}$ and hence a service at the only queue in the third series, which yields a (second) departure from that queue and we have $d_1^{(3)}(2) = 2$. Thus,

$$Q(2) = \begin{bmatrix} 1 & 0 \\ & 1 \end{bmatrix} \quad \text{and} \quad D(2) = \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 0 \\ & & 2 \end{bmatrix}.$$

The corresponding tableau $\tau(2)$ is

1	3
---	---

Similarly, at time 3, $a_3 = 1$, and we get

$$Q(3) = \begin{bmatrix} 2 & 0 \\ & 2 \end{bmatrix} \quad \text{and} \quad D(3) = \begin{bmatrix} 2 & 0 & 0 \\ & 2 & 0 \\ & & 3 \end{bmatrix}.$$

The corresponding tableau $\tau(3)$ is

1	1	3
---	---	---

At time 4, $a_4 = 2$, and there is a departure from the second queue in the first series. This provides a service at the second queue in the second series, which is nonempty; so we have a departure from that queue as well. Thus,

$$Q(4) = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \quad \text{and} \quad D(4) = \begin{bmatrix} 2 & 1 & 0 \\ & 2 & 1 \\ & & 3 \end{bmatrix}.$$

The corresponding tableau $\tau(4)$ is

1	1	3
2		

At time 5, there is a service at the third queue in the first series, and a customer departs. That's all. Thus,

$$Q(5) = \begin{bmatrix} 1 & 0 \\ & 1 \end{bmatrix} \quad \text{and} \quad D(5) = \begin{bmatrix} 2 & 1 & 1 \\ & 2 & 1 \\ & & 3 \end{bmatrix}.$$

The corresponding tableau $\tau(5)$ is

1	1	3
2		
3		

At time 6, there is a service at the second queue in the first series, and a customer departs; this yields a service at the second queue in the second series and a customer departs from there also. Thus,

$$Q(6) = \begin{bmatrix} 0 & 1 \\ & 0 \end{bmatrix} \quad \text{and} \quad D(6) = \begin{bmatrix} 2 & 2 & 1 \\ & 2 & 2 \\ & & 3 \end{bmatrix}.$$

The corresponding tableau $\tau(6)$ is

1	1	3
2	2	
3		

At time 7, there is a service at the second queue in the first series, but this queue is empty and it is not used; this yields a service at the *first* queue in the second series, leading to a departure from that queue and consequently a service at (and departure from) the first (and only) queue in the third series. Thus,

$$Q(7) = \begin{bmatrix} 0 & 1 \\ & 1 \end{bmatrix} \quad \text{and} \quad D(7) = \begin{bmatrix} 2 & 2 & 1 \\ & 3 & 2 \\ & & 4 \end{bmatrix}.$$

The final tableau $\tau(7)$ is

1	1	2	3
2	2		
3			

5. RANDOM WALK IN A WEYL CHAMBER

In this section we record some properties of the conditioned walk of Theorem 1.1, and extend its definition beyond the case $p_1 < \cdots < p_k$.

The random walk X is a Markov chain on \mathbb{Z}_+^k with $X(0) = o$ and transition matrix

$$P(x, y) = p^{y-x} 1_{\{y-x \in b\}}.$$

Denote by P_x the law of the walk started at $x \in W \cap \mathbb{Z}_+^k$.

We will refer to the random walk with $p_1 = \cdots = p_k$ as the *homogeneous walk*.

Denote by s_l the Schur polynomial associated with the integer partition $l_1 \geq l_2 \geq \cdots \geq l_k \geq 0$.

Lemma 5.1. *For any $r \in \mathcal{P}_k$, the function $h_r : \mathbb{Z}_+^k \rightarrow \mathbb{R}$, defined by*

$$h_r(x) = p^{-x} s_{x^*}(r) 1_{x \in W},$$

is harmonic for P . Note that h_r is strictly positive on $W \cap \mathbb{Z}_+^k$.

Proof. This follows immediately from the identity

$$\sum_i s_{l+e_i}(r) = s_l(r),$$

which in turn can be seen as a special case of the Weyl character formula, or can be verified directly using the formula

$$s_l(p) = \det \left(p_i^{l_j + k - j} \right) / \det \left(p_i^{k-j} \right).$$

□

Lemma 5.2. *Suppose $0 < p_1 < \cdots < p_k$. Then, for any $x \in W \cap \mathbb{Z}_+^k$,*

$$P_x(X(n) \in W, \text{ for all } n \geq 0) = C p^{-x} s_{x^*}(p),$$

where C is a constant independent of x . In particular, the transition matrix associated with the conditioned walk of Theorem 1.1 is given by

$$(30) \quad \hat{P}(x, y) = \frac{p^{-y} s_{y^*}(p)}{p^{-x} s_{x^*}(p)} P(x, y) = \frac{s_{y^*}(p)}{s_{x^*}(p)} 1_{\{y-x \in b\}}.$$

Proof. When $r = p$, the Doob transform of the random walk X via the harmonic function h_r has transition matrix \hat{P} . Note that this can also be regarded as the Doob transform of the homogeneous walk via the function $x \mapsto s_{x^*}(kp)$. It follows from the asymptotic analysis of the Green function associated with the (Poissonized) homogeneous walk presented in [28] that, if $\kappa(x, y)$ is the Martin kernel associated with the homogeneous walk, then

$$\kappa(x, y) \rightarrow \text{constant} \times s_{x^*}(kp)$$

whenever y tends to infinity in W in the direction p . Thus, by standard Doob-Hunt theory (see, for example, [15], [43]), any realisation of the corresponding Doob transform, starting from the origin o , almost surely goes to infinity in the direction p . Moreover, any Doob transform on W which almost surely goes to infinity in the direction p is necessarily the same Doob transform. It therefore suffices to show that the “properly” conditioned walk of Theorem 1.1, which is the Doob transform of X via the harmonic function

$$g(x) = P_x(X(n) \in W, \text{ for all } n \geq 0),$$

almost surely goes to infinity in the direction p . But this follows immediately from the estimate, denoting the law of the conditioned walk by \hat{P} ,

$$\hat{P}(|X(n) - pn| > \epsilon n) \leq P(|X(n) - pn| > \epsilon n)/g(x) \leq Ke^{-c(\epsilon)n}/g(x),$$

where $c(\epsilon) > 0$, and a standard Borel-Cantelli argument. □

Note that the transition matrix \hat{P} is well-defined by (30) for any $p \in \mathcal{P}_k$ and, by the symmetry of the Schur polynomials, is symmetric in the p_i .

The proof of Lemma 5.2 given above is presented in more detail in [34], where an explicit formula for the constant C is also given.

6. THE ROBINSON-SCHENSTED ALGORITHM WITH RANDOM WORDS

Having made the connection between the path-transformation $G^{(k)}$ and the Robinson-Schensted algorithm, we will now give a direct proof of Theorem 1.1, purely in the latter context. In fact, we will present a more general result, which does not require the condition $p_1 < \dots < p_k$.

Let ξ_1, ξ_2, \dots be a sequence of independent random variables with common distribution $p \in \mathcal{P}_k$. Let $(S(n), T(n))$ be the pair of semistandard and standard tableaux associated, by the Robinson-Schensted correspondence (with column-insertion), with the random word $\xi_1 \xi_2 \dots \xi_n$. Here, $T(n)$ is the recording tableau. Denote the shape of $S(n)$ by $\lambda(n)$; the weight (or type) of $S(n)$ is $X(n)$, where

$$X_i(n) = |\{1 \leq m \leq n : \xi_m = i\}|.$$

Note that X is the random walk discussed throughout this paper, with transition matrix

$$P(x, y) = p^{y-x} 1_{\{y-x \in b\}}.$$

The joint law of $(S(n), T(n))$ is given, for $\text{sh } \sigma = \text{sh } \tau \vdash n$, by

$$(31) \qquad P(S(n) = \sigma, T(n) = \tau) = p^\sigma,$$

and, for $x \in C = \{x \in \mathbb{Z}_+^k : x_1 \geq \dots \geq x_k\}$,

$$P(\lambda(n) = x) = s_x(p) f_x.$$

Here p^σ is shorthand for p^a , where a is the weight of the tableau σ , and f_x is the number of standard tableaux with shape x . The formula (31) follows immediately from the fact that the Robinson-Schensted correspondence with column-insertion, as in the case with row-insertion, is bijective.

Consider the Doob transform of P on C , defined by its transition matrix

$$(32) \quad Q(x, y) = \frac{p^{-y}s_y(p)}{p^{-x}s_x(p)}P(x, y) = \frac{s_y(p)}{s_x(p)}1_{\{y-x \in b\}}.$$

Theorem 6.1. λ is a Markov chain on C with transition matrix Q .

Proof. For $x, y \in C$, we will write $x \nearrow y$ if $y - x \in b$. Recall that a standard tableau τ with entries $\{1, 2, \dots, n\}$ can be identified with a sequence of integer partitions

$$l(1) \nearrow l(2) \nearrow \dots \nearrow l(n),$$

where $l(m)$ is the shape of the subtableau of τ consisting only of the entries $\{1, 2, \dots, m\}$. Since $T(n)$ is a recording tableau, it is identified in this way with the sequence

$$\lambda(1) \nearrow \lambda(2) \nearrow \dots \nearrow \lambda(n).$$

Thus, summing (31) over semistandard tableaux σ with a given shape $l(n) \vdash n$, we obtain

$$(33) \quad P(\lambda(1) = l(1), \dots, \lambda(n) = l(n)) = \sum_{\text{sh } \sigma = l(n)} p^\sigma = s_{l(n)}(p),$$

and so, for $x \nearrow y \vdash n + 1$,

$$\begin{aligned} &P(\lambda(n+1) = y \mid \lambda(1) = l(1), \dots, \lambda(n-1) = l(n-1), \lambda(n) = x) \\ &= \frac{P(\lambda(1) = l(1), \dots, \lambda(n-1) = l(n-1), \lambda(n) = x, \lambda(n+1) = y)}{P(\lambda(1) = l(1), \dots, \lambda(n-1) = l(n-1), \lambda(n) = x)} \\ &= \frac{s_y(p)}{s_x(p)}, \end{aligned}$$

as required. \square

Recalling the connection between $G^{(k)}$ and the Robinson-Schensted algorithm described in the previous section, and comparing Q with the transition matrix \hat{P} defined by (30), we deduce the following generalisation of Theorem 1.1.

Corollary 6.2. $\hat{X} = G^{(k)}(X)$ is a Markov chain on $W \cap \mathbb{Z}_+^k$ with transition matrix \hat{P} .

We will now record two lemmas which will yield an explicit description of the joint law of X and \hat{X} , and an intertwining relationship between their respective transition matrices.

Denote by κ_{xy} the number of tableaux of shape x and weight y . (These are the Kostka numbers.)

Lemma 6.3.

$$(34) \quad P(X(n) = y \mid \lambda(m), m \leq n) = K(\lambda(n), y),$$

where

$$(35) \quad K(x, y) = \frac{p^y}{s_x(p)} \kappa_{xy}.$$

Proof. First note that the σ -algebra generated by $\{\lambda(m), m \leq n\}$ is precisely the same as the σ -algebra generated by $T(n)$. Thus, the conditional law of $X(n)$, given $\{\lambda(m), m \leq n\}$, is the same as the conditional law of $X(n)$, given $T(n)$. But this only depends on the shape, $\lambda(n)$, of $T(n)$, and is given by

$$K(x, y) := P(X(n) = y \mid \lambda(n) = x) = \frac{p^y}{s_x(p)} \kappa_{xy},$$

as required. \square

We will now show that P and Q are intertwined via the Markov kernel K ; that is,

Lemma 6.4. $QK = KP$.

Proof. We have

$$\begin{aligned} (QK)(x, z) &= \sum_{y \in C} Q(x, y) K(y, z) \\ &= \sum_{y \in C} P(x, y) \frac{p^{-y} s_y(p)}{p^{-x} s_x(p)} \frac{p^z}{s_y(p)} \kappa_{yz} \\ &= \sum_i p_i p^x p^{-x-e_i} \frac{p^z}{s_x(p)} \kappa_{x+e_i, z} \\ &= \frac{p^z}{s_x(p)} \sum_i \kappa_{x+e_i, z}. \end{aligned}$$

On the other hand,

$$\begin{aligned} (KP)(x, z) &= \sum_y K(x, y) P(y, z) \\ &= \sum_i K(x, z - e_i) p_i \\ &= \sum_i p_i \frac{p^{z-e_i}}{s_x(p)} \kappa_{x, z-e_i} \\ &= \frac{p^z}{s_x(p)} \sum_i \kappa_{x, z-e_i}. \end{aligned}$$

The statement of the lemma now follows from the identity

$$\sum_i \kappa_{x+e_i, z} = \sum_i \kappa_{x, z-e_i};$$

to see that this holds, observe that, for any $q \in \mathbb{R}_+^k$,

$$\sum_z q^z \sum_i \kappa_{x+e_i, z} = |q| s_x(q) = \sum_z q^z \sum_i \kappa_{x, z-e_i}.$$

\square

Corollary 6.5. If we set $J(x, y) = K(x^*, y)$, then

$$(36) \quad P(X(n) = y \mid \hat{X}(m), m \leq n, \hat{X}(n) = x) = J(x, y)$$

and $\hat{P}J = JP$.

For a discussion on the role of intertwining in the context of Pitman's $2M - X$ theorem (the case $k = 2$), see [38].

It is instructive to note that Theorem 6.1 also follows from Lemmas 6.3 and 6.4, provided we can show that there is a class of functions of the form $K\varphi$, $\varphi : \mathbb{Z}_+^k \rightarrow \mathbb{R}$, which separate probability distributions on C . Indeed, by Lemmas 6.3 and 6.4,

$$\begin{aligned} & E[(K\varphi)(\lambda(n+1)) | \lambda(m), m \leq n] \\ &= E[E[\varphi(X(n+1)) | \lambda(n+1)] | \lambda(m), m \leq n] \\ &= E[E[\varphi(X(n+1)) | \lambda(m), m \leq n+1] | \lambda(m), m \leq n] \\ &= E[\varphi(X(n+1)) | \lambda(m), m \leq n] \\ &= \sum_y K(\lambda(n), y) E[\varphi(X(n+1)) | X(n) = y] \\ &= \sum_y K(\lambda(n), y) \sum_z P(y, z) \varphi(z) \\ &= [(KP)\varphi](\lambda(n)) \\ &= [(QK)\varphi](\lambda(n)) \\ &= [Q(K\varphi)](\lambda(n)), \end{aligned}$$

which would imply that λ is a Markov chain with transition matrix Q if the functions $K\varphi$ were determining. To find such a class of functions, we recall that the matrix

$$\{\kappa_{xy}, (x, y) \in C^2\}$$

is invertible (see, for example, [31]). Thus, if we set, for $q \in \mathbb{R}_+^k$,

$$(37) \quad \varphi_q(y) = p^{-y} \sum_{z \in C} \kappa_{yz}^{(-1)} q^z s_z(p) 1_{\{y \in C\}},$$

we have $(K\varphi_q)(x) = q^x$, and these functions are clearly determining.

By exactly the same arguments as those given in the proof of Theorem 6.1, if $\mu(n)$ denotes the shape of the tableau obtained by applying the Robinson-Schensted algorithm with row insertion to the random word $\xi_1 \cdots \xi_n$, we obtain

Theorem 6.6. μ is a Markov chain on C with transition matrix Q .

7. POISSONIZED VERSION

We will now define a continuous version of $G^{(k)}$, and state Poissonized versions of Corollaries 6.2 and 6.5. This is an interesting setting in its own right, since the conditioned walk in this case is closely related to the Charlier ensemble and process (see, for example, [27], [28]), but more importantly it provides a convenient framework in which to apply Donsker's theorem and obtain the Brownian analogue of Corollary 6.2, as was presented in [36] in the case $p = (1/k, \dots, 1/k)$. Moreover, given the connection we have now made with the Robinson-Schensted algorithm, this continuous path-transformation can also be regarded as a continuous analogue of the Robinson-Schensted algorithm (see section 10 for further remarks in this direction).

Let $D_0(\mathbb{R}_+)$ denote the space of cadlag paths $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ with $f(0) = 0$. We will extend the definition of the operations \triangle and ∇ to a continuous context. For $f, g \in D_0(\mathbb{R}_+)$, define $f \triangle g \in D_0(\mathbb{R}_+)$ and $f \nabla g \in D_0(\mathbb{R}_+)$ by

$$(38) \quad (f \triangle g)(t) = \inf_{0 \leq s \leq t} [f(s) + g(t) - g(s)]$$

and

$$(39) \qquad (f \nabla g)(t) = \sup_{0 \leq s \leq t} [f(s) + g(t) - g(s)].$$

As in the discrete case, these operations are not associative: unless otherwise delineated by parentheses, the default order of operations is from left to right; for example, when we write $f \triangle g \triangle h$, we mean $(f \triangle g) \triangle h$.

Define a sequence of mappings $\Gamma^{(k)} : D_0(\mathbb{R}_+)^k \rightarrow D_0(\mathbb{R}_+)^k$ by

$$(40) \qquad \Gamma^{(2)}(f, g) = (f \triangle g, g \nabla f),$$

and, for $k > 2$,

$$(41) \qquad \begin{aligned} \Gamma^{(k)}(f_1, \dots, f_k) &= (f_1 \triangle f_2 \triangle \dots \triangle f_k, \\ \Gamma^{(k-1)}(f_2 \nabla f_1, f_3 \nabla (f_1 \triangle f_2), \dots, f_k \nabla (f_1 \triangle \dots \triangle f_{k-1}))). \end{aligned}$$

Let $N = (N_1, \dots, N_k)$ be a continuous-time random walk with generator $Gf(x) = \sum_i \mu_i [f(x + e_i) - f(x)]$. Denote by \mathbb{R}_x the law of N started from x , and by (R_t) the corresponding semigroup. For convenience, we will also denote the corresponding transition kernel by $R_t(x, y)$. Set $p_i = \mu_i / |\mu|$, and denote by \mathbb{S}_x the law of the h_p -transform on W , started at x , and by S_t the corresponding semigroup.

Note that the embedded discrete-time random walk in N has the same law as X . That is, if $\tau_n = \inf\{t \geq 0 : |N(t)| = n\}$ and $Y(n) = N(\tau_n)$, then Y is a random walk on \mathbb{Z}_+^k with transition matrix P .

Theorem 7.1. *The law of $M = \Gamma^{(k)}(N)$ under \mathbb{R}_o is the same as the law of N under \mathbb{S}_o .*

Moreover, if J is the Markov kernel defined in the previous section, then

Theorem 7.2.

$$(42) \qquad \mathbb{R}_o(N(t) = y \mid M(s), s \leq t) = J(M(t), y),$$

and for $t \geq 0$ we have $S_t J = J R_t$.

8. BROWNIAN MOTION IN A WEYL CHAMBER AND RANDOM MATRICES

Let X be a standard Brownian motion in \mathbb{R}^d , and let \mathbb{P}_x denote the law of X started at x . Denote the corresponding semigroup and transition kernel by (P_t) , and the natural filtration of X by (\mathcal{F}_t) .

Recall that

$$W = \{x \in \mathbb{R}^d : x_1 \leq \dots \leq x_d\}$$

and denote by \tilde{P}_t the semigroup of the process killed at the first exit time

$$T = \inf\{t \geq 0 : X(t) \notin W\}.$$

Define \mathbb{Q}_x , for $x \in W^\circ$, by

$$\mathbb{Q}_x|_{\mathcal{F}_t} = \frac{h(X(t \wedge T))}{h(x)} \cdot \mathbb{P}_x|_{\mathcal{F}_t},$$

where h is the Vandermonde function $h(x) = \prod_{i < j} (x_j - x_i)$. Denote the corresponding semigroup by Q_t .

The measure

$$\mathbb{Q}_o = \lim_{W^\circ \ni x \rightarrow 0} \mathbb{Q}_x$$

is well-defined, and can be interpreted as the law of the eigenvalue-process associated with Hermitian Brownian motion [17], [20]. The law of $X(1)$ under \mathbb{Q}_o is the familiar Gaussian Unitary Ensemble (GUE) of random matrix theory.

In [36] it was shown, by applying Donsker's theorem in the context of Theorem 7.1 with $\mu = (1, \dots, 1)$, that

Theorem 8.1. *The law of $\Gamma^{(d)}(X)$ under \mathbb{P}_o is the same as the law of X under \mathbb{Q}_o .*

In particular,¹ $(X_d \nabla \cdots \nabla X_1)(1)$ has the same law as the largest eigenvalue of a $d \times d$ GUE random matrix; this had been observed earlier by Baryshnikov [5] and by Gravner, Tracy and Widom [21]. A similar representation was obtained in [11].

Here we record some additional properties of the process $R = \Gamma^{(d)}(X)$, and its relationship with X , which are inherited, in the same application of Donsker's theorem, from Theorem 7.2.

Theorem 8.2.

$$(43) \quad \mathbb{P}_o(X(t) \in dx \mid R(s), s \leq t; R(t) = r) = L(r, dx),$$

where L is characterised by

$$(44) \quad \int_{\mathbb{R}^k} e^{\lambda \cdot y} L(x, dy) = \frac{\det(e^{\lambda_i x_j})}{h(\lambda)h(x)} =: c_\lambda(x).$$

Also, for $t \geq 0$ we have $Q_t L = L P_t$.

The intertwining $Q_t L = L P_t$ can also be seen as a direct consequence of the Harish-Chandra/Itzykson-Zuber formula [22], [25] for the Laplace transform of the conditional law of the diagonal of a GUE random matrix given its eigenvalues, using the fact that the diagonal of a Hermitian Brownian motion evolves according to the semigroup P_t and the eigenvalues evolve according to the semigroup Q_t . It is also easily verified, using the Karlin-MacGregor formula, that

$$\tilde{P}_t(x, y) = \sum_{\sigma \in S_d} \text{sgn}(\sigma) P_t(x, \sigma y).$$

We will now present analogous results for Brownian motion with drift. Fix $\mu \in \mathbb{R}^d$, and denote by $\mathbb{P}_x^{(\mu)}$ the law of Brownian motion in \mathbb{R}^d with drift μ . Denote by $(P_t^{(\mu)})$ the corresponding semigroup and by $(\tilde{P}_t^{(\mu)})$ the semigroup of the process killed at the first exit time T of the Weyl chamber W . Define h_μ by

$$(45) \quad h_\mu(x) = e^{-\mu \cdot x} |\det(e^{\mu_i x_j})|.$$

It is easy to check directly that h_μ is a positive harmonic function for $\tilde{P}_t^{(\mu)}$. Define

$$\mathbb{Q}_x^{(\mu)} \Big|_{\mathcal{F}_t} = \frac{h_\mu(X(t \wedge T))}{h_\mu(x)} \cdot \mathbb{P}_x^{(\mu)} \Big|_{\mathcal{F}_t}.$$

Denote the corresponding semigroup by $Q_t^{(\mu)}$. Recalling the absolute continuity relationship

$$(46) \quad \tilde{P}_t^{(\mu)}(x, y) = e^{\mu \cdot (y-x) - \|\mu\|_2^2 t/2} \tilde{P}_t(x, y),$$

¹See Remark 2(ii) in section 10 below.

we can write

$$(47) \quad Q_t^{(\mu)}(x, y) = \frac{h_\mu(y)}{h_\mu(x)} \tilde{P}_t^{(\mu)}(x, y) = \frac{\det(e^{\mu_i y_j})}{\det(e^{\mu_i x_j})} e^{-\|\mu\|_2^2 t/2} \tilde{P}_t(x, y),$$

and we note that this is symmetric in the μ_i .

It is easy to verify that the measure

$$Q_o^{(\mu)} = \lim_{W^\circ \ni x \rightarrow 0} Q_x^{(\mu)}$$

is well-defined.

Applying Donsker's theorem in the context of Theorem 7.1, as in [36], we obtain

Theorem 8.3. *The law of $\Gamma^{(d)}(X)$ under $\mathbb{P}_o^{(\mu)}$ is the same as the law of X under $Q_o^{(\mu)}$.*

As in the discrete case, we remark that the law $Q_o^{(\mu)}$ is symmetric in the drifts μ_i .

We also have, by the same application of Donsker's theorem, the following analogue of Theorem 7.2.

Theorem 8.4.

$$(48) \quad \mathbb{P}_o^{(\mu)}(X(t) \in dx \mid R(s), s \leq t; R(t) = r) = L^{(\mu)}(r, dx),$$

where

$$(49) \quad L^{(\mu)}(x, dy) = c_\mu(x)^{-1} e^{\mu \cdot y} L(x, dy).$$

Also, for $t \geq 0$,

$$(50) \quad Q_t^{(\mu)} L^{(\mu)} = L^{(\mu)} P_t^{(\mu)}.$$

The intertwining relationship (50) can also be verified directly using $Q_t L = L P_t$ and (46).

For related work on reflecting Brownian motions and non-colliding diffusions see [7], [11], [12], [13], [16], [23], [35], [39] and references therein.

9. AN APPLICATION IN QUEUEING THEORY

In this section, using the connection with the Robinson-Schensted correspondence obtained in section 3, we will write down a formula for the "transient distribution" of a series of $M/M/1$ queues in tandem. There are many papers on this topic for the case of a single queue, where the solution is given in terms of modified Bessel functions; see [2] and references therein. In [3], the case of two queues was considered and a solution obtained, but the techniques used there do not seem to extend easily to higher dimensions.

Consider a series of $M/M/1$ queues in tandem, k in number, driven by Poisson processes N_1, \dots, N_k with respective intensities μ_1, \dots, μ_k . The first queue has infinitely many customers, and the remaining queues are initially empty. At every point of N_i , there is a service at the i^{th} queue and, provided that queue is not empty, a customer departs and joins the $(i + 1)^{th}$ queue (or leaves the system if $i = k$).

Denote by $D(t) = (D_1(t), \dots, D_k(t))$ the respective numbers of customers to depart from each queue up to time t . Note that, since there are always infinitely many customers in the first queue, D_1 is a Poisson process; we can thus ignore the first queue, think of the second queue as the first in a series of $k - 1$ queues, and

think of D_1 as the arrival process at the first of these $k - 1$ queues in the series. This is a more conventional setup in queueing theory. The state of the system is described by the queue lengths

$$(51) \quad Q_1 = D_1 - D_2, \dots, Q_{k-1} = D_{k-1} - D_k.$$

We will write down a formula for the law of $D(t)$ which, in turn, yields the law of $Q(t) = (Q_1(t), \dots, Q_{k-1}(t))$.

Without loss of generality, we can assume that $|\mu| = 1$. The de-Poissonized version of this problem is to consider the usual random walk X , with $p = \mu$, and consider the law of $\delta(n) = (D^{(k)}(X))(n)$. But we know this law, from sections 3 and 6. It is the law of $\beta(S(n))$, where $S(n)$ of the random semistandard tableau obtained when one applies the Robinson-Schensted algorithm with column-insertion to the random word $\xi_1 \cdots \xi_k$ and $\beta_i(\tau)$ denotes the number of i 's in the i^{th} row of a tableau τ .

Thus,

$$(52) \quad P(D(t) = d) = e^{-t} \sum_{n \geq 0} \frac{t^n}{n!} P(\delta(n) = d),$$

where

$$(53) \quad P(\delta(n) = d) = \sum_{l \geq d, l \vdash n} \sum_{\tau \vdash l} p^\tau f_l 1_{\{\beta(\tau) = d\}}.$$

This formula is complicated in general, but simplifies in certain cases.

Consider the case $k = 2$. In this case, we have only one summand:

$$(54) \quad P(\delta(n) = d) = p_1^{d_1} p_2^{n-d_1} f_{(n-d_2, d_2)}.$$

By the hook-length formula (see, for example, [18]), for $n \geq d_1 + d_2$,

$$f_{(n-d_2, d_2)} = n! \frac{n - 2d_2 + 1}{(n - d_2 + 1)! d_2!}.$$

Thus, recalling that $p = \mu$,

$$(55) \quad P(D(t) = d) = e^{-t} \sum_{n \geq d_1 + d_2} t^n \mu_1^{d_1} \mu_2^{n-d_1} \frac{n - 2d_2 + 1}{(n - d_2 + 1)! d_2!}.$$

It follows that

$$(56) \quad P(Q(t) = q) = (\mu_1/\mu_2)^q e^{-t} \sum_{m \geq q} (m+1) (\mu_2 t)^m I_{m+1}(2\sqrt{\mu_1 \mu_2} t).$$

Let $s_{l/d}$ denote the Schur polynomial associated with the skew-tableau l/d (see, for example, [18]). We will use the following formula:

$$(57) \quad \sum_l s_{l/d}(x) \frac{t^{|l|} f_l}{|l|!} = e^{|x|t} \frac{t^{|d|} f_d}{|d|!}.$$

This follows from the identity

$$(58) \quad \sum_l s_{l/d}(x) s_l(y) = s_d(y) \prod_{i,j} (1 - x_i y_j)^{-1}$$

(this is a variant of Cauchy's identity; see, for example, [31, pp. 62–70]) and the fact that

$$(59) \quad \lim_{n \rightarrow \infty} s_l \left(\frac{t}{n} \cdot 1^n \right) = \frac{t^{|l|} f_l}{|l|!}.$$

In the case $k = 3$, if $p_2 = p_3$, we have

$$(60) \quad \begin{aligned} P(D_1(t) = d_1, D_2 \geq d_2, D_3(t) = d_3) &= e^{-t} p^d \sum_l s_{l/d}(p_2, p_3) \frac{t^{|l|} f_l}{|l|!} \\ &= e^{-p_1 t} p^d \frac{t^{|d|} f_d}{|d|!}. \end{aligned}$$

It would be interesting to compare (60) with the explicit formulas obtained in [3] for this case.

In the general case, we can simplify the formula (52) if d is constant. Suppose $d_i = m$ for all i . Then, using (57) and the hook-length formula,

$$(61) \quad P(D(t) = d) = e^{-t} p^d \sum_l s_{l/d}(p_2, p_3, \dots, p_k) \frac{t^{|l|} f_l}{|l|!}$$

$$(62) \quad = e^{-p_1 t} p^d \frac{t^{|d|} f_d}{|d|!}$$

$$(63) \quad = e^{-p_1 t} (t^k p_1 p_2 \cdots p_k)^m \prod_{i \leq k} \frac{\Gamma(i + m)}{\Gamma(i)}.$$

It follows that

$$(64) \quad P(Q_1(t) = Q_2(t) = \cdots = Q_{k-1}(t) = 0) = e^{-p_1 t} H(t^k p_1 p_2 \cdots p_k),$$

where

$$(65) \quad H(s) = \sum_{m \geq 0} s^m \prod_{i \leq k} \frac{\Gamma(i)}{\Gamma(i + m)}.$$

Finally we remark that, by Theorem 6.1 and the symmetry of the Schur polynomials, the law of the process D_k is symmetric in the parameters p_1, \dots, p_k .

10. CONCLUDING REMARKS

1. *The Krawchouk process:* A binomial version of Theorem 1.1 was presented in [28]. This states that, if X is a random walk in Z_+^k with transition matrix

$$P(x, y) = C p^{y-x} 1_{y-x \in \{0,1\}^k}$$

(C is a normalising constant), then, assuming $p_1 < \cdots < p_k$ and extending slightly the domain of $G^{(k)}$, we see that $G^{(k)}(X)$ has the same law as that of X conditioned to stay forever in W . In this case, a similar connection can be made with the dual Robinson-Schensted-Knuth (RSK) correspondence for zero-one matrices, and analogues of all the main results of section 6 can be obtained similarly. The Schur polynomials again play an important role. See [34] for details.

2. *Properties of $\Gamma^{(k)}$:* The following continuous analogues of Lemmas 2.1–2.4 can be readily verified. Denote by $C_0(\mathbb{R}_+, \mathbb{R}^k)$ the set of continuous functions $f: \mathbb{R}_+ \rightarrow \mathbb{R}^k$ with $f(0) = o$. For $f \in C_0(\mathbb{R}_+, \mathbb{R}^k)$, where $k \geq 2$,

$$(i) \quad |\Gamma^{(k)}(f)| = |f|,$$

- (ii) $\Gamma_k^{(k)}(f) = f_k \nabla \cdots \nabla f_1$,
- (iii) $f(t) = [\Gamma^{(k)}(f)](t) + \Phi^{(k)}([\Gamma^{(k)}(f)](t, u), u \geq t)$, where $\Phi^{(k)}$ is defined on a suitable domain as the continuous analogue of $F^{(k)}$.

In this continuous setting, the identity (ii) can be verified directly using the “sup-integration by parts” formula

$$\sup_{0 < s < t} \left\{ \sup_{0 < r < s} u(r) + v(s) \right\} \bigvee \sup_{0 < s < t} \left\{ u(s) + \sup_{0 < r < s} v(r) \right\} = \sup_{0 < s < t} u(s) + \sup_{0 < s < t} v(s),$$

for $u, v \in C_0(\mathbb{R}_+, \mathbb{R}^k)$. This is a “max-plus” analogue of the usual integration by parts formula and is easily verified by the method of Laplace.

3. *A continuous Robinson-Schensted algorithm:* Given the connection we have made between the path-transformations $G^{(k)}$ and the Robinson-Schensted algorithm, the mappings $\Gamma^{(k)}$ can be used to define a continuous version of the Robinson-Schensted algorithm. More precisely, let \mathbb{C}_{GC} denote the *Gelfand-Cetlin cone*, which consists of triangular arrays of real numbers

$$(66) \quad (x_j^{(i)}, 1 \leq i \leq k, 1 \leq j \leq i)$$

satisfying $x_j^{(i)} \geq x_j^{(i-1)} \geq x_{j+1}^{(i)}$, for all i, j . Points in the Gelfand-Cetlin cone can be regarded as continuous analogues of semistandard tableaux. The continuous analogue of a word is a continuous function $f : [0, 1] \rightarrow \mathbb{R}^k$ with $f(0) = o$: denote the set of these functions by $C_0([0, 1], \mathbb{R}^k)$. Define a map $\phi : C_0([0, 1], \mathbb{R}^k) \rightarrow \mathbb{C}_{GC}$ as follows. For convenience, let $\Gamma^{(1)}$ be the identity transformation. If we set $x = \phi(f_1, \dots, f_k)$, then, for each $1 \leq i \leq k$,

$$(67) \quad x^{(i)} = ([\Gamma_i^{(i)}(f_1, \dots, f_i)](1), \dots, [\Gamma_1^{(i)}(f_1, \dots, f_i)](1)).$$

The continuous analogue of the corresponding “recording tableau” is the path

$$(68) \quad \rho(f) = \{[\Gamma^{(k)}(f)](t), 0 \leq t \leq 1\} \in C_0([0, 1], W).$$

By analogy with the discrete Robinson-Schensted algorithm, the function f can be uniquely recovered from the pair $\phi(f)$ and $\rho(f)$. A more detailed discussion on the properties of this continuous Robinson-Schensted algorithm will be presented elsewhere.

4. *GUE minors:* Let A be a $k \times k$ GUE random matrix, and denote the eigenvalues of the i^{th} minor $(A_{lm}, l, m \leq i)$ by $\lambda_1^{(i)} \geq \cdots \geq \lambda_i^{(i)}$, for $i \leq k$. In the above context, Baryshnikov [5] showed that, if $(B(t), 0 \leq t \leq 1)$ is a standard Brownian motion in \mathbb{R}^k , then the random vector

$$(\phi_1^{(1)}(B), \dots, \phi_1^{(k)}(B))$$

has the same law as

$$(\lambda_1^{(1)}, \dots, \lambda_1^{(k)}).$$

In [5], Donsker’s theorem is applied in the context of a random semistandard tableau with the same law as $T(n)$ of section 6 in the homogeneous case $p_1 = \cdots = p_k$. We can thus extend Baryshnikov’s arguments using the representation for the Robinson-Schensted algorithm given in this paper and the continuity of the mappings $\Gamma^{(i)}$, $i \leq k$, to see that, in fact, $\phi(B)$ has the same law as

$$(\lambda^{(1)}, \dots, \lambda^{(k)}).$$

However, it is easy to see,² by considering the case $k = 2$, that these identities do not extend to the process level (that is, with ϕ defined simultaneously on intervals $[0, t]$ instead of just $[0, 1]$ and “GUE” replaced by “Hermitian Brownian motion”).

5. *Related topics:* The intertwining of Lemma 6.4 is closely related to the work of Biane on quantum random walks [8], [9], [10]. The Robinson-Schensted correspondence is of fundamental importance to the representation theory of S_n and $GL(n)$. Related topics in representation theory (which are certainly connected to results presented in this paper) include Littelmann's path model for the finite-dimensional representations of $GL(n)$ (see, for example, [30]), crystal bases, and representations of quantum groups (see, for example, [14], [29]).

APPENDIX

Proof of Lemma 2.1. The first identity is trivial:

$$\begin{aligned} (y \triangledown x)(n) &= \max_{0 \leq m \leq n} [y(m) + x(n) - x(m)] \\ &= x(n) + y(n) + \max_{0 \leq m \leq n} [y(m) - y(n) - x(m)] \\ &= x(n) + y(n) - (x \triangle y)(n). \end{aligned}$$

If we set $z = y - x$, and $s(n) = \max_{0 \leq m \leq n} z(m)$, then the second identity is equivalent to the well-known fact that

$$s(n) = \min_{l \leq n} [2s(l) - z(l)].$$

□

Proof of Lemma 2.2. Fix $x \in \Lambda_k$, and write $G^{(k)} = G^{(k)}(x)$ unless otherwise indicated (similarly for $D^{(k)}$ and $T^{(k)}$). We will first show that $|G^{(k)}| = |x|$. We will prove this by induction on k . The case $k = 2$ is given by Lemma 2.1. Assume the induction hypothesis for $k - 1$. We recall from the definitions that

$$(69) \quad G^{(k)} = \left(D_k^{(k)}, G^{(k-1)} \left(T^{(k)} \right) \right).$$

By the induction hypothesis,

$$(70) \quad |G^{(k)}| = D_k^{(k)} + |T^{(k)}|.$$

Recall that, for $i \geq 2$,

$$(71) \quad D_i^{(k)} = D_{i-1}^{(k)} \triangle x_i$$

and

$$(72) \quad T_{i-1}^{(k)} = x_i \triangledown D_{i-1}^{(k)}.$$

Thus, by Lemma 2.1,

$$(73) \quad x_i = D_i^{(k)} - D_{i-1}^{(k)} + T_{i-1}^{(k)}$$

for each $i \geq 2$; summing this over i and recalling that $D_1^{(k)} = x_1$ yields

$$(74) \quad |x| = D_k^{(k)} + |T^{(k)}|;$$

so we are done.

²Bougerol and Jeulin, private communication

We will now show that

$$(75) \quad x(n) = G^{(k)}(n) + F^{(k)} \left(G^{(k)}(l) - G^{(k)}(n), l \geq n \right),$$

for some function $F^{(k)}$ to be defined. Again we will prove this by induction on k , and note that for $k = 2$, this is given by Lemma 2.1.

A recursive definition of $F^{(k)}$ will be implicit in the induction argument. Recall that

$$(76) \quad G^{(k)} = \left(D_k^{(k)}, G^{(k-1)} \left(T^{(k)} \right) \right).$$

Assuming the induction hypothesis for $k - 1$, we have, for $i \geq 2$,

$$(77) \quad T_{i-1}^{(k)}(n) = G_i^{(k)}(n) + F^{(k-1)} \left((G_2^{(k)}, \dots, G_k^{(k)})(n, l) \right).$$

Thus, for $i \geq 2$, using (73) and the fact that

$$D_i^{(k)}(n) - D_{i-1}^{(k)}(n) = \max_{l \geq n} [D_i^{(k)}(n, l) - T_{i-1}^{(k)}(n, l)],$$

we have

$$(78) \quad x_i(n) = G_i^{(k)}(n) + J_i^{(k)} \left((D_i^{(k)}, G_2^{(k)}, \dots, G_k^{(k)})(n, l) \right),$$

where $J_i^{(k)}$ is defined on a suitable domain. It is important to note here that $J_i^{(k)}$ does not depend on n .

In this way, recalling that $D_k^{(k)} = G_1^{(k)}$, we obtain

$$(79) \quad x_k(n) = G_k^{(k)}(n) + F_k^{(k)} \left(G^{(k)}(l) - G^{(k)}(n), l \geq n \right),$$

where the function $F_k^{(k)}$ is implicitly defined by this identity (on a suitable domain) and does not depend on n . Observe that we can also recover the sequence of future increments $x_k(l) - x_k(n)$ as a function, which does not depend on n , of the sequence $\{G^{(k)}(l) - G^{(k)}(n), l \geq n\}$.

We can now recover the values $x_{k-1}(n)$, $x_{k-2}(n)$, and so on, as follows. By equations (73) with $i = k$,

$$(80) \quad D_{k-1}^{(k)}(n) = G_1^{(k)}(n) - x_k(n) + T_{k-1}^{(k)}(n).$$

It follows that the sequence $\{D_{k-1}^{(k)}(n, l), l \geq n\}$ is a function, which does not depend on n , of the sequence $\{G^{(k)}(l) - G^{(k)}(n), l \geq n\}$. Combining this with (78), we see that

$$(81) \quad x_{k-1}(n) = G_{k-1}^{(k)}(n) + F_{k-1}^{(k)} \left(G^{(k)}(l) - G^{(k)}(n), l \geq n \right),$$

where $F_{k-1}^{(k)}$ is implicitly defined by this identity (on a suitable domain) and does not depend on n . Similarly, we can recover the sequence of future increments $x_k(l) - x_k(n)$ as a function, which does not depend on n , of the sequence $\{G^{(k)}(l) - G^{(k)}(n), l \geq n\}$, and so on. Finally, $x_1(n)$ is obtained using $|x| = |G^{(k)}|$. \square

Proof of Lemma 2.3. We want to show that, for $(a, b, c) \in \Lambda_3$,

$$(82) \quad a \nabla (c \triangle b) \nabla (b \nabla c) = a \nabla b \nabla c,$$

and for $(w, x, y) \in \Lambda_3$,

$$(83) \quad w \triangle (y \nabla x) \triangle (x \triangle y) = w \triangle x \triangle y.$$

First note that these identities are equivalent. To see this, set $a(n) = n - w(n)$, $b(n) = n - x(n)$ and $c(n) = n - y(n)$. Then plug these into (82) to obtain (83). We will therefore restrict our attention to the identity (83).

Let $d = x \triangle y$, $t = y \nabla x$, $q = x - d$ and $u = y - d$. Then (83) becomes

$$(84) \quad w \triangle (x + u) \triangle (y - u) = w \triangle x \triangle y.$$

That is, the output of a series of queues in tandem driven by $(w, x + u, y - u)$ is the same as that of the series driven by (w, x, y) . Set

$$\begin{aligned} d_1 &= w \triangle x, \\ d_2 &= w \triangle x \triangle y, \\ \tilde{d}_1 &= w \triangle (x + u), \\ \tilde{d}_2 &= w \triangle (x + u) \triangle (y - u), \end{aligned}$$

and

$$\begin{aligned} q_1 &= w - d_1, \\ q_2 &= d_1 - d_2, \\ \tilde{q}_1 &= w - \tilde{d}_1, \\ \tilde{q}_2 &= \tilde{d}_1 - \tilde{d}_2. \end{aligned}$$

We want to show that $d_2 = \tilde{d}_2$. From the above definitions, this is equivalent to showing that

$$q_1(n) + q_2(n) = \tilde{q}_1(n) + \tilde{q}_2(n)$$

for all $n \geq 0$. We will prove this by induction on n .

The induction hypothesis H is:

$q_1 + q_2 = \tilde{q}_1 + \tilde{q}_2$, and *either*

- (i) $\tilde{q}_2 - q_2 \geq 0$ and $q - q_2 = 0$, or
- (ii) $\tilde{q}_2 - q_2 = 0$ and $q - q_2 \geq 0$.

When $n = 0$ we have $q = q_1 = q_2 = \tilde{q}_1 = \tilde{q}_2 = 0$, and the induction hypothesis is trivially satisfied. Assume the induction hypothesis holds at time $n - 1$. Note that $(w, x, y - u, u) \in \Lambda_4$; that is, only one of these quantities, if any, can increase by one at time n . We will consider the following five cases, which are exhaustive and mutually exclusive, separately.

- (a) $(w, x, y - u, u)(n) = (w, x, y - u, u)(n - 1)$,
- (b) $w(n) - w(n - 1) = 1$,
- (c) $x(n) - x(n - 1) = 1$,
- (d) $(y - u)(n) - (y - u)(n - 1) = 1$,
- (e) $u(n) - u(n - 1) = 1$.

Case (a): $(w, x, y - u, u)(n) = (w, x, y - u, u)(n - 1)$. In this case, nothing changes, and so H is preserved.

Case (b): $w(n) - w(n - 1) = 1$. In this case, $q_1(n) = q_1(n - 1) + 1$ and $\tilde{q}_1(n) = \tilde{q}_1(n - 1) + 1$, the other quantities remain unchanged, and H is preserved.

Case (c): $x(n) - x(n - 1) = 1$. Then $q(n) = q(n - 1) + 1$.

Suppose $q_1(n - 1) > \tilde{q}_1(n - 1) > 0$. Then $q_1(n) = q_1(n - 1) - 1$ and $q_2(n) = q_2(n - 1) + 1$. Thus, $q - q_2$ and $q_1 + q_2$ do not change. Also, $\tilde{q}_1(n) = \tilde{q}_1(n - 1) - 1$ and $\tilde{q}_2(n) = \tilde{q}_2(n - 1) + 1$. Thus, $\tilde{q}_2 - q_2$ and $\tilde{q}_1 + \tilde{q}_2$ do not change either; so we still have $q_1 + q_2 = \tilde{q}_1 + \tilde{q}_2$, and H is preserved.

Now suppose $q_1(n-1) > \tilde{q}_1(n-1) = 0$. Note that this implies $\tilde{q}_2(n-1) - q_2(n-1) > 0$, so that we are initially in case (i) of the induction hypothesis. In this case, $q_1(n) = q_1(n-1) - 1$ and $q_2(n) = q_2(n-1) + 1$, but \tilde{q}_1 and \tilde{q}_2 do not change. Thus, $q - q_2$, $q_1 + q_2$ and $\tilde{q}_1 + \tilde{q}_2$ do not change. The quantity $\tilde{q}_2 - q_2$ decreases by one, but remains nonnegative; so we remain in case (i), and H is preserved.

Finally, if $q_1(n-1) = \tilde{q}_1(n-1) = 0$, then we are initially in case (ii) of H . There is no change to q_1 , q_2 , \tilde{q}_1 or \tilde{q}_2 , but q increases by one, and so we remain in case (ii), and H is preserved.

Case (d): $(y - u)(n) - (y - u)(n-1) = 1$. In this case, $q(n-1) > 0$ and q decreases by one. The values of q_1 and \tilde{q}_1 do not change.

If we are in case (i) at time $n-1$, then $\tilde{q}_2(n-1) \geq q_2(n-1) > 0$, and so both q_2 and \tilde{q}_2 also decrease by one; thus, we remain in case (i), and H is preserved.

If we are in case (ii) at time $n-1$, then $\tilde{q}_2(n-1) = q_2(n-1)$, and either q_2 and \tilde{q}_2 both decrease by one or both remain unchanged. Either way, we remain in case (ii), and H is preserved.

Case (e): $u(n) - u(n-1) = 1$. Then $q(n-1) = q_2(n-1) = 0$, and we are initially in case (i) of H . The values of q , q_1 and q_2 will not change. If $\tilde{q}_1 > 0$, then \tilde{q}_1 decreases by one and \tilde{q}_2 increases by one; otherwise, \tilde{q}_1 and \tilde{q}_2 do not change. Either way, we remain in case (i), and H is preserved. \square

Proof of Lemma 2.4. We will prove this by induction on k . It is certainly true for $k = 2$, from the definition of $G^{(2)}$. Recall the definition of $G^{(k)}$,

$$(85) \quad G^{(k)} = \left(D_k^{(k)}, G^{(k-1)} \left(T^{(k)} \right) \right).$$

By the induction hypothesis, that the lemma is true for $G^{(k-1)}$, we have

$$(86) \quad G_k^{(k)} = T_{k-1}^{(k)} \nabla T_{k-2}^{(k)} \nabla \cdots \nabla T_1^{(k)}.$$

We will now repeatedly apply Lemma 2.3:

$$\begin{aligned} T_{k-1}^{(k)} \nabla T_{k-2}^{(k)} &= x_k \nabla D_{k-1}^{(k)} \nabla T_{k-2}^{(k)} \\ &= x_k \nabla (D_{k-2}^{(k)} \triangle x_{k-1}) \nabla (x_{k-1} \nabla D_{k-2}^{(k)}) \\ &= x_k \nabla x_{k-1} \nabla D_{k-2}^{(k)}. \end{aligned}$$

Similarly,

$$\begin{aligned} x_k \nabla x_{k-1} \nabla D_{k-2}^{(k)} \nabla T_{k-3}^{(k)} &= x_k \nabla x_{k-1} \nabla (D_{k-3}^{(k)} \triangle x_{k-2}) \nabla (x_{k-2} \nabla D_{k-3}^{(k)}) \\ &= x_k \nabla x_{k-1} \nabla x_{k-2} \nabla D_{k-3}^{(k)}, \end{aligned}$$

and so on. \square

REFERENCES

- [1] F. Baccelli, G. Cohen, G.J. Olsder and J.-P. Quadrat. *Synchronization and Linearity: An Algebra for Discrete Event Systems*. Wiley, 1992. MR **94b**:93001
- [2] F. Baccelli and W.A. Massey. A sample path analysis of the M/M/1 queue. *J. Appl. Probab.* 26 (1989) 418–422. MR **90i**:60065
- [3] F. Baccelli and W.A. Massey. *A transient analysis of the two-node series Jackson network*. INRIA Rapport de Recherche No. 852 (1988).
- [4] J. Baik, P. Deift and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* 12 (1999), no. 4, 1119–1178. MR **2000e**:05006

- [5] Yu. Baryshnikov. GUES and queues. *Probab. Theory Related Fields* 119 (2001) 256–274. MR **2002a**:60165
- [6] A. Berenstein and A.N. Kirillov. The Robinson-Schensted-Knuth bijection, quantum matrices and piecewise linear combinatorics. To appear in: *Proceedings of the 13th International Conference on Formal Power Series and Algebraic Combinatorics, Arizona, 2001*.
- [7] Ph. Biane. Quelques propriétés du mouvement brownien dans un cône. *Stoch. Proc. Appl.* 53 (1994), no. 2, 233–240. MR **95j**:60129
- [8] Ph. Biane. Théorème de Ney-Spitzer sur le dual de $SU(2)$. *Trans. Amer. Math. Soc.* 345, no. 1 (1994) 179–194. MR **95a**:60103
- [9] Ph. Biane. *Intertwining of Markov semi-groups, some examples*. Séminaire de Probabilités, XXIX, 30–36, Lecture Notes in Math., 1613, Springer-Verlag, Berlin, 1995. MR **98k**:46117
- [10] Ph. Biane. Quantum random walk on the dual of $SU(n)$. *Probab. Theory Related Fields* 89 (1991), no. 1, 117–129. MR **93a**:46119
- [11] Ph. Bougerol and Th. Jeulin. Paths in Weyl chambers and random matrices. *Probab. Theory Related Fields* 124 (2002) 517–543.
- [12] K. Burdzy. A three-dimensional Brownian path reflected on a Brownian path is a free Brownian path. *Letters Math. Phys.* 27 (1993) 239–241. MR **94c**:60132
- [13] K. Burdzy and D. Nualart. Brownian motion reflected on Brownian motion. *Probab. Theory Related Fields*, 122 (2002), 471–493.
- [14] E. Date, M. Jimbo and T. Miwa. *Representations of $U_q(\mathfrak{gl}(n, \mathbb{C}))$ at $q = 0$ and the Robinson-Schensted correspondence*. Physics and Mathematics of Strings, 185–211, World Sci. Publishing, Teaneck, NJ, 1990. MR **92h**:17012
- [15] J.L. Doob. *Classical Potential Theory and its Probabilistic Counterpart*. Springer-Verlag, 1984. MR **85k**:31001
- [16] Y. Doumerc. *An asymptotic link between LUE and GUE and its spectral interpretation*. Preprint.
- [17] F.J. Dyson. A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* 3 (1962) 1191–1198. MR **26**:5904
- [18] William Fulton. *Young Tableaux*. London Mathematical Society student texts: 35. Cambridge University Press, 1997. MR **99f**:05119
- [19] P.W. Glynn and W. Whitt. Departures from many queues in series. *Ann. Appl. Prob.* 1 (1991), no. 4, 546–572. MR **92i**:60162
- [20] D. Grabiner. Brownian motion in a Weyl chamber, non-colliding particles, and random matrices. *Ann. IHP Probab. Statist.* 35 (1999), no. 2, 177–204. MR **2000i**:60091
- [21] J. Gravner, C.A. Tracy and H. Widom. Limit theorems for height fluctuations in a class of discrete space and time growth models. *J. Statist. Phys.* 102 (2001), 1085–1132. MR **2002d**:82065
- [22] Harish-Chandra. Invariant differential operators on a semisimple Lie algebra. *Proc. Nat. Acad. Sci. U.S.A.* 42 (1956), 252–253. MR **18**:218c
- [23] J. M. Harrison and R.J. Williams. On the quasireversibility of a multiclass Brownian service station. *Ann. Probab.* 18 (1990) 1249–1268. MR **91i**:60204
- [24] A.R. Its, C.A. Tracy and H. Widom. Random words, Toeplitz determinants, and integrable systems. In: *Random Matrices and their Applications*, MSRI Publications, Volume 40, 2001, 245–258. MR **2002i**:82040
- [25] C. Itzykson and J.B. Zuber. The planar approximation. II. *J. Math. Phys.* 21 (1980), no. 3, 411–421. MR **81a**:81068
- [26] K. Johansson. Shape fluctuations and random matrices. *Commun. Math. Phys.* 209 (2000) 437–476. MR **2001h**:60177
- [27] K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure, *Ann. Math.* (2) 153 (2001), no. 1, 259–296. MR **2002g**:05188
- [28] Wolfgang König, Neil O'Connell and Sebastien Roch. Non-colliding random walks, tandem queues and the discrete ensembles. *Elect. J. Probab.* 7 (2002) Paper no. 5, 1–24.
- [29] Alain Lascoux, Bernard Leclerc, and Jean-Yves Thibon. The plactic monoid. In: *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.
- [30] Peter Littelmann. The path model, the quantum Frobenius map and standard monomial theory. *Algebraic groups and their representations (Cambridge, 1997)*, 175–212, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 517, Kluwer Acad. Publ., Dordrecht, 1998. MR **99m**:20096

- [31] I.G. MacDonald. *Symmetric Functions and Hall Polynomials*. Second edition, Oxford, 1995. MR **96h**:05207
- [32] M.L. Mehta. *Random Matrices: Second Edition*. Academic Press, 1991. MR **92f**:82002
- [33] Neil O'Connell. *Random matrices, non-colliding processes and queues*. Lecture Notes in Math., vol. 1801, Springer-Verlag, 2003.
- [34] Neil O'Connell. Conditioned random walks and the RSK correspondence. *J. Phys. A* 36 (2003) 3049-3066.
- [35] Neil O'Connell and Marc Yor. Brownian analogues of Burke's theorem. *Stoch. Proc. Appl.* 96 (2) (2001) pp. 285-304. MR **2002h**:60175
- [36] Neil O'Connell and Marc Yor. A representation for non-colliding random walks. *Elect. Commun. Probab.* 7 (2002) 1-12.
- [37] J.W. Pitman. One-dimensional Brownian motion and the three-dimensional Bessel process. *Adv. Appl. Probab.* 7 (1975) 511-526. MR **51**:11677
- [38] L.C.G. Rogers and J.W. Pitman. Markov functions. *Ann. Probab.* 9 (1981) 573-582. MR **82j**:60133
- [39] F. Soucaliuc, B. Toth and W. Werner. Reflection and coalescence between independent one-dimensional Brownian paths. *Ann. Inst. Henri Poincaré* 36 (2000) 509-545. MR **2002a**:60139
- [40] R.P. Stanley. *Enumerative Combinatorics*, Volume 2, Cambridge University Press, 1999. MR **2000k**:05026
- [41] C.A. Tracy and H. Widom. Fredholm determinants, differential equations and matrix models. *Comm. Math. Phys.* 163 (1994), no. 1, 33-72. MR **95e**:82005
- [42] C.A. Tracy and H. Widom. On the distributions of the lengths of the longest monotone subsequences in random words. *Probab. Theory Related Fields* 119 (2001), no. 3, 350-380. MR **2002a**:60013
- [43] David Williams. *Diffusions, Markov Processes and Martingales. Volume 1: Foundations*. Wiley, 1979. MR **80i**:60100
- [44] David Williams. Path decomposition and continuity of local time for one-dimensional diffusions I. *Proc. London Math. Soc.* 28 (1974), no. 3, 738-768. MR **50**:3373

MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY CV4 7AL, UNITED KINGDOM
E-mail address: noc@maths.warwick.ac.uk

ON THE IWASAWA λ -INVARIANTS OF REAL ABELIAN FIELDS

TAKAE TSUJI

ABSTRACT. For a prime number p and a number field k , let A_∞ denote the projective limit of the p -parts of the ideal class groups of the intermediate fields of the cyclotomic \mathbb{Z}_p -extension over k . It is conjectured that A_∞ is finite if k is totally real. When p is an odd prime and k is a real abelian field, we give a criterion for the conjecture, which is a generalization of results of Ichimura and Sumida. Furthermore, in a special case where p divides the degree of k , we also obtain a rather simple criterion.

1. INTRODUCTION

Let p be an odd prime number. For a number field k , let k_∞ denote the cyclotomic \mathbb{Z}_p -extension of k with its n -th layer k_n ($n \geq 0$). We denote by A_∞ the projective limit of the p -Sylow subgroup of the ideal class group of each k_n with respect to the relative norms. If k is a totally real number field, it is conjectured that A_∞ is a finite abelian group ([7], [12, p. 316]), which is often called Greenberg's conjecture.

When k is a real abelian field whose degree is not divisible by p , Ichimura and Sumida ([8], [9], [10]) discovered a good method for verifying the conjecture. In this case, we can decompose A_∞ into a direct sum of its χ -parts A_∞^χ for Dirichlet characters χ corresponding to k . Then they gave a necessary and sufficient condition for A_∞^χ to be finite in terms of certain cyclotomic units and some polynomials related to the Kubota-Leopoldt p -adic L -function $L_p(s, \chi)$ associated to χ . It is suitable for a practical computer calculation and, for example, using it they showed that the conjecture holds when $p = 3$ and $k = \mathbb{Q}(\sqrt{m})$ with $1 < m < 10^4$. (Similar criteria are also obtained by [13] and [14].)

In this paper, we study the case where k is a real abelian field of arbitrary degree, including the case $p \mid [k : \mathbb{Q}]$. Although we cannot necessarily decompose A_∞ into direct summands by using characters χ in this case, the conjecture for k follows from the finiteness of all (suitably defined) χ -parts A_∞^χ of A_∞ (cf. Lemma 2.1). The finiteness of A_∞^χ is also conjectured, and the purpose of this paper is to give some criteria for this conjecture for χ to be true.

First, we will generalize the criterion of Ichimura and Sumida to arbitrary characters, especially characters of order divisible by p (Theorem 2.6). In their proof, Ichimura and Sumida ([8], [10]) essentially used a theorem of Iwasawa [11] and Gillard [6] which describes the Galois module structure of "semi-local units modulo

Received by the editors October 27, 2002.

2000 *Mathematics Subject Classification.* Primary 11R23.

Key words and phrases. Iwasawa theory, Greenberg's conjecture, abelian fields.

cyclotomic units" $\mathcal{U}^\chi/\mathcal{C}^\chi$ in terms of the p -adic L -function $L_p(s, \chi)$. Iwasawa and Gillard dealt with only the case where the order of χ is not divisible by p , and this is a main reason why Ichimura and Sumida imposed the same assumption. The author [18] removed the assumption on the order of χ in the theorem of Iwasawa and Gillard and determined the structure of $\mathcal{U}^\chi/\mathcal{C}^\chi$ for arbitrary χ (Fact 2). By using this result, we can prove our theorem by the same way as the proof of Ichimura and Sumida. When the order of χ is p and $\chi(p) = 1$, Fukuda and Komatsu [5] obtained a similar criterion.

In a special case where the order of χ is divisible by p , we also obtain the following simple sufficient condition for the conjecture to be true:

Theorem (Corollary 2.4). *Assume that $\chi\omega^{-1}(p)$ is a nontrivial p -power root of unity, where ω is the Teichmüller character. If the distinguished polynomial $P_\chi(T)$ associated to the p -adic L -function $L_p(s, \chi)$ is irreducible, then A_∞^χ is finite.*

The assumption in the above theorem is satisfied only when the order of χ is divisible by p , since the order of ω is prime to p . In particular, this case was not dealt with in the papers [8], [9] and [10] of Ichimura and Sumida. We will prove the above theorem by using the fact proved in [18] that $\mathcal{U}^\chi/\mathcal{C}^\chi$ is not necessarily a cyclic module.

We remark that there indeed exist characters χ satisfying the condition in the above theorem. We give some numerical examples in §6. We also verify in §6 the conjecture for several real abelian fields by applying our theorem, a generalized criterion of Ichimura and Sumida, to some characters χ .

2. MAIN RESULTS

We fix an odd prime number p . Let χ be a nontrivial $\overline{\mathbb{Q}_p}$ -valued *even* Dirichlet character of the first kind, i.e., the conductor of χ is not divisible by p^2 , and let $k = k_\chi$ be the fixed field of the kernel of χ . We denote by k_∞ the cyclotomic \mathbb{Z}_p -extension of k with its n -th layer k_n ($n \geq 0$). We write $A_n = A_{k_n}$ for the p -Sylow subgroup of the ideal class group of k_n and put $A_\infty = A_{k_\infty} := \varprojlim A_n$, the projective limit being taken with respect to the relative norms. We put $\Delta := \text{Gal}(k/\mathbb{Q})$ and $\Gamma := \text{Gal}(k_\infty/k)$; so $\text{Gal}(k_\infty/\mathbb{Q}) = \Delta \times \Gamma$, since χ is of the first kind. Then we regard A_∞ as a module over the completed group ring $\mathbb{Z}_p[\Delta][\Gamma]$. It is well known that A_∞ is finitely generated and torsion over $\mathbb{Z}_p[\Delta][\Gamma]$ ([12, Theorem 5]). Let \mathcal{O} denote the ring generated by the values of χ over \mathbb{Z}_p . For any $\mathbb{Z}_p[\Delta]$ -module M , we define an \mathcal{O} -module M^χ , the χ -part of M , by

$$M^\chi := \{m \in M \otimes_{\mathbb{Z}_p} \mathcal{O} \mid \delta m = \chi(\delta)m \ \forall \delta \in \Delta\}$$

(for the properties of the χ -part, cf. e.g. [17, §II.1] and [18, §2]). If χ' is a \mathbb{Q}_p -conjugate of χ , then $M^{\chi'} \cong M^\chi$. Then \mathcal{O} -modules A_n^χ are defined ($0 \leq n \leq \infty$) and $A_\infty^\chi = \varprojlim A_n^\chi$ becomes an $\mathcal{O}[\Gamma]$ -module. Let f be the prime-to- p part of the conductor of χ and put $q = fp$. Denote by μ_{p^n} the group of p^n -th roots of unity for $n \geq 0$ and $\mu_{p^\infty} := \bigcup_n \mu_{p^n}$. Identifying Γ with $\text{Gal}(k(\mu_{p^\infty})/k(\mu_p))$ in a natural way, we choose and fix the topological generator γ of Γ such that $\zeta^\gamma = \zeta^{1+q}$ for all $\zeta \in \mu_{p^\infty}$. We identify, as usual, the completed group ring $\mathcal{O}[\Gamma]$ with the power series ring $\Lambda := \mathcal{O}[[T]]$ by $\gamma = 1 + T$. Thus, A_∞^χ is regarded as a module over Λ , and is finitely generated and torsion over Λ . For a finitely generated torsion Λ -module M , denote by $\text{char}_\Lambda M$ the characteristic polynomial of M , which is a uniquely

determined distinguished polynomial times a power of a fixed prime element of \mathcal{O} . We denote by λ_χ (resp. μ_χ) the λ -invariant (resp. μ -invariant) of A_∞^χ . We know that $\mu_\chi = 0$ by the theorem of Ferrero and Washington [4]. Greenberg's conjecture for χ is as follows:

Conjecture. Let χ be an even Dirichlet character of the first kind. It is conjectured that A_∞^χ is finite, that is, $\text{char}_\Lambda A_\infty^\chi = 1$ or $\lambda_\chi = 0$.

Greenberg's conjecture for a real abelian field implies the above. Indeed, by using the theorem of Ferrero and Washington [4], one can show the following:

Lemma 2.1. *Let K be a real abelian field. Then A_{K_∞} is finite if and only if $\lambda_\chi = 0$ for all representatives of \mathbb{Q}_p -conjugate classes of Dirichlet characters χ of the first kind with $k_\chi \subset K_\infty$.*

In our main results of this paper, we give two criteria for the conjecture for χ . To state these, we need to recall the relation between $\text{char}_\Lambda A_\infty^\chi$ and the Kubota-Leopoldt p -adic L -function $L_p(s, \chi)$ associated to χ . By Iwasawa, there exists a unique power series $g_\chi(T)$ in $\mathcal{O}[[T]]$ such that

$$g_\chi((1+q)^s - 1) = L_p(1-s, \chi)$$

(cf. [20, Theorem 7.10]). Using the p -adic Weierstrass preparation theorem and the theorem of Ferrero and Washington [4], one can uniquely write

$$(2.1) \quad g_\chi(T) = u_\chi(T)P_\chi(T)$$

for a distinguished polynomial P_χ in $\mathcal{O}[T]$ and a unit $u_\chi(T)$ of Λ . Put $\lambda_\chi^* = \deg P_\chi(T)$. It follows from the Iwasawa main conjecture proved in [16] that

$$(2.2) \quad \text{char}_\Lambda A_\infty^\chi \mid P_\chi(T),$$

and hence $\lambda_\chi \leq \lambda_\chi^*$ (see (3.2) and Fact 1 in §3). Therefore we have $\text{char}_\Lambda A_\infty^\chi = 1$ if and only if $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$ for all distinguished irreducible factors $P(T)$ of $P_\chi(T)$.

One of our main theorems of this paper is as follows:

Definition. For a distinguished polynomial $P(T)$ in $\mathcal{O}[T]$ such that $P(T) \mid P_\chi(T)$, we define a condition (T) as follows:

$$(T) \quad v_{\mathcal{O}}(Q(q)) < v_{\mathcal{O}}(1 - \chi\omega^{-1}(p)) \quad \text{where} \quad Q(T) = P_\chi(T)/P(T),$$

where $v_{\mathcal{O}}$ denotes an additive valuation of \mathcal{O} and ω the Teichmüller character.

Then we have

Theorem 2.2. *Let $P(T)$ be a distinguished polynomial in $\mathcal{O}[T]$ such that $P(T) \mid P_\chi(T)$. Assume that $P(T)$ satisfies (T) . Then we have $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$.*

For the condition (T) , we can easily see the following:

Lemma 2.3.

- (i) If $\chi\omega^{-1}(p) \notin \mu_{p^\infty}$, there exists no factor of $P_\chi(T)$ satisfying (T) . In particular, if the order of χ is prime to p and $\chi\omega^{-1}(p) \neq 1$, there exists no factor of $P_\chi(T)$ satisfying (T) .
- (ii) If $\chi\omega^{-1}(p) \neq 1$, a factor $P(T)$ of $P_\chi(T)$ satisfies (T) if and only if

$$v_{\mathcal{O}}(P(q)) > v_{\mathcal{O}}(B_{1, \chi\omega^{-1}})$$

holds, where $B_{1, \chi\omega^{-1}}$ denotes the first generalized Bernoulli number.

- (iii) If $\chi\omega^{-1}(p) \in \mu_{p^\infty}$, then $P_\chi(T)$ satisfies (T).
- (iv) ([3, Proposition 2]) If $\chi\omega^{-1}(p) = 1$, then $T - q$ divides $P_\chi(T)$ and is the unique irreducible polynomial satisfying (T).

In the case where $\chi\omega^{-1}(p) = 1$, by (iv) of the above lemma, Theorem 2.2 just asserts that

$$T - q \mid P_\chi(T) \quad \text{and} \quad T - q \nmid \text{char}_\Lambda A_\infty^\chi.$$

This is already proved under the assumption that the order of χ is prime to p ([9, Remark 5]). For other cases, we obtain the following.

Corollary 2.4. Assume that $\chi\omega^{-1}(p) \in \mu_{p^\infty} \setminus \{1\}$. Then $P_\chi(T) \nmid \text{char}_\Lambda A_\infty^\chi$. Furthermore, if $P_\chi(T)$ is irreducible, or if $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$, we have $\lambda_\chi = 0$.

We will give in §6 some examples satisfying the conditions in the above corollary.

Remark 1. For an arbitrary even Dirichlet character χ of the first kind, we can show that $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$ implies $A_n^\chi = \{1\}$ for all $n \geq 0$ (see §5).

Next, we state the other main theorem of this paper, which is a generalization of the result of Ichimura and Sumida [8], [10] to arbitrary characters. We have to prepare some notation. We fix a distinguished polynomial $P(T)$ in $\mathcal{O}[T]$ such that $P(T) \mid P_\chi(T)$. Put $\omega_n = \omega_n(T) = (1+T)^{p^n} - 1$ and $\nu_n = \nu_n(T) = \omega_n/T$ for $n \geq 0$. By the Leopoldt conjecture for p and k_n proved in [1], $\Lambda/(P, \omega_n)$ and $\Lambda/(P, \nu_n)$ are finite abelian groups for any $n \geq 0$. We denote by $m_{P,n}$ the exponent of $\Lambda/(P, \omega_n)$ (resp. $\Lambda/(P, \nu_n)$) if $\chi(p) \neq 1$ (resp. $\chi(p) = 1$). Then we take a polynomial $X_{P,n}(T)$ in $\mathcal{O}[T]$ satisfying

$$(2.3) \quad X_{P,n}(T)P(T) \equiv m_{P,n} \pmod{\begin{cases} \omega_n & \text{if } \chi(p) \neq 1, \\ \nu_n & \text{if } \chi(p) = 1. \end{cases}}$$

This polynomial $X_{P,n}$ is uniquely determined modulo ω_n (resp. ν_n) since ω_n and $P(T)$ are relatively prime. Choose an element $\tilde{Y}_{P,n}$ in $\mathbb{Z}[\Delta][T]$ such that

$$\tilde{Y}_{P,n} \equiv \tilde{X}_{P,n} \pmod{m_{P,n}},$$

where $\tilde{X}_{P,n}$ is an element of $\mathbb{Z}_p[\Delta][T]$ satisfying $\chi(\tilde{X}_{P,n}) = X_{P,n}$. Here we regard χ as a $\mathbb{Z}_p[T]$ -linear homomorphism $\mathbb{Z}_p[\Delta][T] \rightarrow \mathcal{O}[T]$ induced by χ . Let Δ_p (resp. Δ') denote the p -Sylow subgroup (resp. the prime-to- p part) of Δ :

$$\Delta = \Delta_p \times \Delta'.$$

We put $\psi = \chi|_{\Delta'}$. Let

$$e_\psi := \frac{1}{\#\Delta'} \sum_{\delta \in \Delta'} \text{Tr}(\psi(\delta))\delta^{-1}$$

be the idempotent of $\mathbb{Z}_p[\Delta']$ corresponding to ψ , where Tr is the trace map from the field generated by the values of ψ over \mathbb{Q}_p to \mathbb{Q}_p . Let $\alpha \in \mathbb{Z}[\Delta]$ denote an element of Δ_p of order p (resp. $\alpha = 0$) if Δ_p is nontrivial (resp. trivial), and $\mathbf{e}_\chi = \mathbf{e}_{\chi,P,n}$ an element of $\mathbb{Z}[\Delta]$ such that

$$\mathbf{e}_\chi \equiv e_\psi(1 - \alpha) \pmod{m_{P,n}}.$$

For any $m \geq 1$, we fix a primitive m -th root ζ_m of unity with the property that $\zeta_{mm'}^{m'} = \zeta_m$ for all $m' \geq 1$. We use cyclotomic elements in k_n defined as follows:

$$c_n = N_{\mathbb{Q}(\zeta_{fp^{n+1}})/k_n} (1 - \zeta_{fp^{n+1}})^{t-1},$$

where t denotes the cardinality of the residue class field of a prime ideal of k over p . If $f \neq 1$, i.e., the conductor of χ is not equal to p , then the element c_n is a (cyclotomic) unit of k_n . Since the case where $f = 1$ is treated in [8], [9] and [10], we assume $f \neq 1$ in all that follows. By the identification $\gamma = 1 + T$, the element $\tilde{Y}_{P,n}$ of $\mathbb{Z}[\Delta][T]$ can act on the group k_n^\times . For each $n \geq 0$ consider the following condition:

$$(H_{P,n}) \quad c_n^{\mathbf{e}_\chi \tilde{Y}_{P,n}} \notin (k_n^\times)^{m_{P,n}}.$$

We can see that the condition $(H_{P,n})$ does not depend on the choices of $X_{P,n}$, $\tilde{X}_{P,n}$ and $\tilde{Y}_{P,n}$. We remark that, in the case where $\chi(p) = 1$, the condition $(H_{P,0})$ does not hold since $c_0 = 1$ and $m_{P,0} = 1$ for any $P(T)$. The following lemma can be proved in a way similar to [9, Lemma 1] by using Lemma 4.1.

Lemma 2.5. *The condition $(H_{P,n})$ implies $(H_{P,n+1})$.*

Our main theorem is stated as follows:

Theorem 2.6. *Let $P(T)$ be a distinguished polynomial in $\mathcal{O}[T]$ such that $P(T) \mid P_\chi(T)$. Then we have $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$ if and only if the condition $(H_{P,n})$ holds for some $n \geq 0$.*

The above theorem coincides with theorems in [8] and [10] if the exponent of Δ divides $p - 1$.

By Theorems 2.2, 2.6 and (2.2), we obtain

Corollary 2.7. *We have $\lambda_\chi = 0$ if and only if for any distinguished irreducible polynomial $P(T)$ such that $P(T) \mid P_\chi(T)$ and $P(T)$ does not satisfy (T) , the condition $(H_{P,n})$ holds for some $n \geq 0$.*

By the Chebotarev density theorem, we obtain the following:

Corollary 2.8. *We have $\lambda_\chi = 0$ if and only if for any distinguished irreducible factor $P(T)$ of $P_\chi(T)$ that does not satisfy (T) , there exist an integer $n \geq 0$ and a prime ideal \mathfrak{l} of k_n of degree one for which the condition*

$$c_n^{\mathbf{e}_\chi \tilde{Y}_{P,n}} \bmod \mathfrak{l} \notin ((\mathbb{Z}/l\mathbb{Z})^\times)^{m_{P,n}}$$

holds. Here $l\mathbb{Z} = \mathfrak{l} \cap \mathbb{Q}$.

3. PRELIMINARIES

We first recall the Iwasawa main conjecture and show its consequence (2.2). Let M/k_∞ be the maximal abelian p -extension unramified outside p , and let L/k_∞ be the maximal unramified abelian p -extension. As usual, we consider $\text{Gal}(M/k_\infty)$, $\text{Gal}(L/k_\infty)$ and $\text{Gal}(M/L)$ as $\mathbb{Z}_p[\Delta][\Gamma]$ -modules. Let \mathfrak{p} be a prime ideal of k over p . There exists a unique prime ideal \mathfrak{p}_n of k_n over \mathfrak{p} since k is of the first kind. We denote by $U_{n,\mathfrak{p}}$ the group of principal units in the completion $k_{n,\mathfrak{p}}$ of k_n at \mathfrak{p}_n . Put

$$\mathcal{U}_n := \prod_{\mathfrak{p} \mid p} U_{n,\mathfrak{p}},$$

where \mathfrak{p} runs over all prime ideals of k over p . Let E'_n be the group of units ϵ of k_n such that $\epsilon \equiv 1 \bmod \mathfrak{p}_n$ for all $\mathfrak{p}_n \mid p$. Let \mathcal{E}_n be the closure of the image of E'_n under the diagonal map $E'_n \rightarrow \mathcal{U}_n$. Put

$$\mathcal{U} := \varprojlim \mathcal{U}_n, \quad \mathcal{E} := \varprojlim \mathcal{E}_n,$$

where the projective limits are taken with respect to the relative norms. We regard \mathcal{U} and \mathcal{E} as modules over $\mathbb{Z}_p[\Delta][\Gamma]$. By class field theory, we have the following isomorphisms of $\mathbb{Z}_p[\Delta][\Gamma]$ -modules:

$$(3.1) \quad A_\infty \cong \text{Gal}(L/k_\infty), \quad \mathcal{U}/\mathcal{E} \cong \text{Gal}(M/L).$$

(For the latter, see [9, Lemma 3].) Put

$$\mathfrak{X} := \text{Gal}(M/k_\infty).$$

It is known that \mathfrak{X} is finitely generated and torsion over $\mathbb{Z}_p[\Delta][\Gamma]$ ([12, Theorem 17]), and we further see that \mathfrak{X} is finitely generated as a \mathbb{Z}_p -module by [4]. So A_∞ and \mathcal{U}/\mathcal{E} are also finitely generated over \mathbb{Z}_p . Hence we have

$$(3.2) \quad \text{char}_\Lambda A_\infty^\chi \cdot \text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi) = \text{char}_\Lambda \mathfrak{X}^\chi$$

(cf. e.g. [18, §2]). The Iwasawa main conjecture proved in [16] asserts the following:

Fact 1. The torsion Λ -module \mathfrak{X}^χ has the characteristic polynomial $P_\chi(T)$:

$$\text{char}_\Lambda \mathfrak{X}^\chi = P_\chi(T).$$

Hence the relation (2.2), $\text{char}_\Lambda A_\infty^\chi \mid P_\chi(T)$, holds. Furthermore, Greenberg's conjecture for (p, χ) is equivalent to the following:

$$\text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi) = P_\chi(T).$$

Next we recall results on the structures of the Λ -modules \mathcal{U}^χ and \mathcal{U}^χ/C^χ in [11], [6] and [18] (Fact 2), which are essentially used in the proof of our main theorems. (C^χ is a group of cyclotomic units defined below.) Since we are assuming $f \neq 1$, $c_n = N_{\mathbb{Q}(\zeta_{fp^{n+1}})/k_n}(1 - \zeta_{fp^{n+1}})^{t-1}$ is a unit in E'_n , and hence $c_n \in \mathcal{E}_n$. We regard c_n as an element of $\mathcal{E}_n \otimes \mathcal{O}$. We define $\xi_\chi \in \mathcal{O}[\Delta]$ by

$$\xi_\chi := \sum_{\delta \in \Delta} \chi(\delta)^{-1} \delta.$$

Then $c_n^{\xi_\chi} = \sum_{\delta \in \Delta} c_n^\delta \otimes \chi(\delta)^{-1}$ is an element of \mathcal{E}_n^χ . We can see that $N_{m,n}(c_m) = c_n$ for all $m \geq n \geq 0$. Then we put

$$c_\infty^{\xi_\chi} := (c_n^{\xi_\chi})_{n \geq 0} \in \mathcal{U}^\chi = \varprojlim \mathcal{U}_n^\chi$$

and denote by C^χ the submodule of \mathcal{U}^χ generated by $c_\infty^{\xi_\chi}$ over Λ .

Let \mathbb{T}_n denote the \mathbb{Z}_p -torsion of \mathcal{U}_n^χ and put $\mathbb{T} := \varprojlim \mathbb{T}_n$, where the projective limit is taken with respect to the relative norms. Then we have

$$\mathbb{T}_n^\chi \cong (\mathcal{O}/(p^{n+1}, 1 - \chi\omega^{-1}(p))) \otimes_{\mathbb{Z}_p} \mathbb{Z}_p(1),$$

and hence

$$(3.3) \quad \mathbb{T}^\chi \cong \begin{cases} \{1\} & \text{if } \chi\omega^{-1}(p) \neq 1, \\ \Lambda/(T - q) & \text{if } \chi\omega^{-1}(p) = 1, \end{cases}$$

where $\mathbb{Z}_p(1) = \varprojlim \mu_{p^n}$.

The following fact plays an important role to prove Theorems 2.2 and 2.6. When the order of χ is not divisible by p , this is the theorem of Iwasawa and Gillard ([11] and [6]).

Fact 2. ([18, Theorem 2.1, Proposition 4.2]) There is a natural Λ -homomorphism

$$\Psi : \mathcal{U}^\chi \longrightarrow \Lambda$$

for which the kernel is \mathbb{T}^χ and the image is $(T - q, 1 - \chi\omega^{-1}(p))$, the ideal of Λ generated by $T - q$ and $1 - \chi\omega^{-1}(p)$. Furthermore, we have

$$\Psi(c_\infty^{\xi_\chi}) = g_\chi(T).$$

Put $\mathcal{V}_n^\chi := \bigcap_{m \geq n} N_{m,n}(\mathcal{U}_m^\chi)$, with the norm maps $N_{m,n}$ from k_m to k_n . In the following lemma, we determine the structure of the Λ -modules \mathcal{V}_n^χ to deal with the Λ -module $\mathcal{U}_n^\chi/\mathcal{E}_n^\chi$.

Lemma 3.1.

(i) The projection $\mathcal{U}^\chi \rightarrow \mathcal{V}_n^\chi$ induces the following isomorphisms:

$$\mathcal{V}_n^\chi \cong \begin{cases} \mathcal{U}^\chi/(\mathcal{U}^\chi)^{\omega_n} & \text{if } \chi(p) \neq 1, \\ \mathcal{U}^\chi/(\mathcal{U}^\chi)^{\nu_n} & \text{if } \chi(p) = 1. \end{cases}$$

(ii) We have an isomorphism $\mathcal{U}_n^\chi/\mathcal{V}_n^\chi \cong \mathcal{O}/(1 - \chi(p))$.

By this lemma, the homomorphism Ψ in Fact 2 induces a Λ -homomorphism Ψ_n as follows:

$$\Psi_n : \mathcal{V}_n^\chi \longrightarrow \Lambda/(\vartheta_n^\chi)$$

where $\vartheta_n^\chi = \omega_n$ (resp. ν_n) if $\chi(p) \neq 1$ (resp. $\chi(p) = 1$). We further see that the kernel of Ψ_n is \mathbb{T}_n^χ , and we can determine the cokernel by Fact 2.

Proof. We prove this lemma in a way similar to [6, Proposition 2] by using local class field theory. Let D be the decomposition group of p in Δ and put $\tilde{\chi} = \chi|_D$. Fix a prime ideal \mathfrak{p} of k over p . Write U_n for $U_{n,\mathfrak{p}}$, the group of the principal units of $k_{n,\mathfrak{p}}$, for all $n \geq 0$. We regard $U := \varprojlim U_n$ and U_n as modules over $\mathbb{Z}_p[D][\Gamma] = \mathbb{Z}_p[D][[T]]$ in a natural way (with $\gamma = 1 + T$). Put $V_n^{\tilde{\chi}} := \bigcap_{m \geq n} N_{m,n}(U_m^{\tilde{\chi}})$ with norm map $N_{m,n}$ from $k_{m,\mathfrak{p}}$ to $k_{n,\mathfrak{p}}$. Then we have Λ -isomorphisms

$$(3.4) \quad \mathcal{U}_n^\chi \cong U_n^{\tilde{\chi}} \otimes_{\mathcal{O}_D} \mathcal{O}, \quad \mathcal{V}_n^\chi \cong V_n^{\tilde{\chi}} \otimes_{\mathcal{O}_D} \mathcal{O}$$

(see [18, §4]). Here \mathcal{O}_D denotes the ring generated by the values of $\tilde{\chi}$ over \mathbb{Z}_p .

If $\chi(p) = 1$, i.e., $\tilde{\chi} = 1$, then $k_{\mathfrak{p}} = \mathbb{Q}_p$. In this case, the assertions (i) and (ii) follow from [6, Proposition 2], [20, Lemma 13.53] and (3.4).

We assume that $\chi(p) \neq 1$, i.e., $\tilde{\chi} \neq 1$. Let \widehat{X}_n be the p -adic completion of $k_{n,\mathfrak{p}}^\times$; $\widehat{X}_n := \varprojlim_m k_{n,\mathfrak{p}}^\times / (k_{n,\mathfrak{p}}^\times)^{p^m}$, and let $\widehat{X}_\infty := \varprojlim \widehat{X}_n$. Since the (local) cyclotomic extension $k_{\infty,\mathfrak{p}}/k_{\mathfrak{p}}$ is totally ramified, we have the following commutative diagram with exact rows of $\mathbb{Z}_p[D][\Gamma]$ -modules:

$$(3.5) \quad \begin{array}{ccccccc} 1 & \longrightarrow & U & \longrightarrow & \widehat{X}_\infty & \longrightarrow & \mathbb{Z}_p \longrightarrow 1 \\ & & \phi \downarrow & & \downarrow & & \parallel \\ 1 & \longrightarrow & U_n & \longrightarrow & \widehat{X}_n & \longrightarrow & \mathbb{Z}_p \longrightarrow 1. \end{array}$$

Here D and Γ act on \mathbb{Z}_p trivially. Let $k_{n,\mathfrak{p}}^{ab}$ be the maximal abelian p -extension of $k_{n,\mathfrak{p}}$ for $0 \leq n \leq \infty$. We have isomorphisms $\widehat{X}_n \cong \text{Gal}(k_{n,\mathfrak{p}}^{ab}/k_{n,\mathfrak{p}})$ and $N\widehat{X}_n := \bigcap_{m \geq n} N_{m,n}(\widehat{X}_m) \cong \text{Gal}(k_{n,\mathfrak{p}}^{ab}/k_{\infty,\mathfrak{p}})$ of $\mathbb{Z}_p[D][\Gamma]$ -modules, by local class field theory, and an isomorphism

$$\text{Gal}(k_{n,\mathfrak{p}}^{ab}/k_{\infty,\mathfrak{p}}) \cong \text{Gal}(k_{\infty,\mathfrak{p}}^{ab}/k_{\infty,\mathfrak{p}})/\omega_n \text{Gal}(k_{\infty,\mathfrak{p}}^{ab}/k_{\infty,\mathfrak{p}})$$

holds. Then we have $N\widehat{X}_n \cong \widehat{X}_\infty/\widehat{X}_\infty^{\omega_n}$. Therefore the kernel of ϕ in (3.5) is $\widehat{X}_\infty^{\omega_n}$. So the kernel of the projection $U^{\tilde{\chi}} \rightarrow V_n^{\tilde{\chi}}$ is $(\widehat{X}_\infty^{\omega_n})^{\tilde{\chi}}$. Clearly we have $(\widehat{X}_\infty^{\omega_n})^{\tilde{\chi}} \supset (\widehat{X}_\infty^{\tilde{\chi}})^{\omega_n}$. Let x^{ω_n} be an element of $(\widehat{X}_\infty^{\omega_n})^{\tilde{\chi}}$, i.e., $x \in \widehat{X}_\infty \otimes_{\mathbb{Z}_p} \mathcal{O}_D$ and $(x^{\omega_n})^{\tilde{\chi}(\delta)-\delta} = 1$ for all $\delta \in D$. Since \widehat{X}_∞ has no nontrivial element that is killed by ω_n ([12, Theorem 25]), we have $x^{\tilde{\chi}(\delta)-\delta} = 1$, that is, $x^{\omega_n} \in (\widehat{X}_\infty^{\tilde{\chi}})^{\omega_n}$. Thus $(\widehat{X}_\infty^{\omega_n})^{\tilde{\chi}} = (\widehat{X}_\infty^{\tilde{\chi}})^{\omega_n}$. We also see that $\widehat{X}_\infty^{\tilde{\chi}} = U^{\tilde{\chi}}$ by (3.5) under the assumption $\tilde{\chi} \neq 1$. Hence, the assertion (i) has been proved.

To prove (ii), we first determine the structure of $(NU_n)^{\tilde{\chi}}/V_n^{\tilde{\chi}}$, where NU_n denotes $\bigcap_{m \geq n} N_{m,n}(U_m)$. If the order of $\tilde{\chi}$ is prime to p , the functor $(*)^{\tilde{\chi}}$ is exact. Therefore, in this case, we have $V_n^{\tilde{\chi}} = (NU_n)^{\tilde{\chi}}$, that is, $(NU_n)^{\tilde{\chi}}/V_n^{\tilde{\chi}} \cong \mathcal{O}_D/(1 - \chi(p))$. Then we consider the case where p divides the order of $\tilde{\chi}$. Decompose $D = D_p \times D'$ with the p -Sylow subgroup D_p of D and put $\tilde{\psi} = \tilde{\chi}|_{D'}$. We use the following lemma.

Lemma 3.2 (cf. [17, Lemma II.2]). *Assume that D_p is nontrivial and that $\tilde{\chi}|_{D_p}$ is a faithful character of D_p . Let C be the subgroup of D_p of order p and N_C the norm of C in $\mathbb{Z}_p[D]$. Then, for a $\mathbb{Z}_p[D]$ -module M , we have a $\mathbb{Z}_p[D]$ -isomorphism*

$$M^{\tilde{\chi}} \cong \ker(N_C : e_{\tilde{\psi}}M \rightarrow e_{\tilde{\psi}}M),$$

where $e_{\tilde{\psi}}$ denotes the idempotent of $\mathbb{Z}_p[D']$ corresponding to $\tilde{\psi}$.

We see that $\widehat{H}^0(C, U) = 0$ since k_p/\mathbb{Q}_p is tamely ramified. Therefore, by the above lemma and the fact that the kernel of $U \rightarrow NU_n$ is \widehat{X}^{ω_n} proved above, we have an isomorphism

$$(NU_n)^{\tilde{\chi}}/V_n^{\tilde{\chi}} \cong \widehat{H}^0(C, (\widehat{X}^{e_{\tilde{\psi}}})^{\omega_n}).$$

The fact that \widehat{X}_∞ has no nontrivial element killed by ω_n implies an isomorphism $\widehat{H}^0(C, (\widehat{X}^{e_{\tilde{\psi}}})^{\omega_n}) \cong \widehat{H}^0(C, \widehat{X}^{e_{\tilde{\psi}}})$. By using $\widehat{H}^i(C, U) = 0$ for $i = -1, 0$ and (3.5), we have isomorphisms $\widehat{H}^0(C, \widehat{X}^{e_{\tilde{\psi}}}) \cong \widehat{H}^0(C, e_{\tilde{\psi}}\mathbb{Z}_p) \cong \mathcal{O}_D/(1 - \chi(p))$. Hence

$$(NU_n)^{\tilde{\chi}}/V_n^{\tilde{\chi}} \cong \mathcal{O}_D/(1 - \chi(p)).$$

We know that $U_n/NU_n \cong \mathbb{Z}_p$ (see [20, Lemma 13.53]). Thus, we have

$$(NU_n)^{\tilde{\chi}} = U_n^{\tilde{\chi}}$$

under the assumption $\tilde{\chi} \neq 1$. Thus, we have completed the proof. \square

Finally, we shall show the freeness of \mathcal{E}^χ (Lemma 3.5). The following is well known.

Fact 3 ([12, Theorem 18]). \mathfrak{X} has no nontrivial finite $\mathbb{Z}_p[\Delta][\Gamma]$ -submodule.

We need the following lemma, which follows from the Leopoldt conjecture for k_n and p proved in [1].

Lemma 3.3 (cf. [20, §5-5]).

(i) *The inclusion $E'_n \rightarrow \mathcal{E}_n$ induces an isomorphism*

$$E'_n/E_n^{p^a} \cong \mathcal{E}_n/\mathcal{E}_n^{p^a}$$

for any $a \geq 0$.

(ii) *\mathcal{E}_n is \mathbb{Z}_p -torsion free.*

By the above lemma, we can regard \mathbb{T} as a submodule of \mathcal{U}/\mathcal{E} and also of \mathfrak{X} . We show the following lemma by using Fact 3.

Lemma 3.4. \mathfrak{X}/\mathbb{T} has no nontrivial finite $\mathbb{Z}_p[\Delta][[\Gamma]]$ -submodule.

Proof. We identify $\mathbb{Z}_p[\Delta][[\Gamma]]$ with $\mathbb{Z}_p[\Delta][[T]]$ by $\gamma = 1 + T$. For a $\mathbb{Z}_p[\Delta][[T]]$ -module M , we put $M_T := \{m \in M \mid Tm = 0\}$. If M is nontrivial of finite order, then M_T is nontrivial. Thus, by Fact 3, it suffices to show that $(\mathfrak{X}/\mathbb{T})_T \cong \mathfrak{X}_T$. Since \mathbb{T}_T is trivial, we have an exact sequence

$$1 \longrightarrow \mathfrak{X}_T \longrightarrow (\mathfrak{X}/\mathbb{T})_T \xrightarrow{\tau} (\mathbb{T} \cap T\mathfrak{X})/T\mathbb{T} \longrightarrow 1,$$

where τ is induced by $x \mapsto Tx$ for $x \in \mathfrak{X}$. Let M_0 be the maximal abelian extension unramified outside p . The restriction map $\mathfrak{X} \rightarrow \text{Gal}(M_0/k)$ induces a surjection $\mathbb{T} \rightarrow \mathbb{T}_0$, and hence an isomorphism

$$\mathbb{T}/(\mathbb{T} \cap T\mathfrak{X}) \cong \mathbb{T}_0.$$

On the other hand, we see that \mathbb{T}_n is trivial or $\mu_p^{n+1} \otimes_{\mathbb{Z}_p} \mathbb{Z}_p[\Delta/D]$, where D denotes the decomposition group of p in Δ for all $n \geq 0$ according to whether \mathbb{T}_0 is trivial or not. Thus we have

$$\mathbb{T}/T\mathbb{T} \cong \mathbb{T}_0.$$

Therefore we have $\mathbb{T} \cap T\mathfrak{X} = T\mathbb{T}$, as desired. \square

Using Lemma 3.4, we prove the following:

Lemma 3.5. $\Psi(\mathcal{E}^\chi)$ is a principal ideal generated by $\text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi)$. In particular, \mathcal{E}^χ is a free Λ -module of rank one.

Proof. There exists an element $f(T)$ of Λ such that the principal ideal $(f(T))$ of Λ has a submodule $\Psi(\mathcal{E}^\chi)$ of finite index. We first show that $f(T)$ is contained in $\Psi(\mathcal{U}^\chi) = (T - q, 1 - \chi\omega^{-1}(p))$. If $\chi\omega^{-1}(p) \notin \mu_{p^\infty} \setminus \{1\}$, this assertion is clear since $\Psi(\mathcal{U}^\chi)$ is a principal ideal of Λ . Assume $\chi\omega^{-1}(p) \in \mu_{p^\infty} \setminus \{1\}$. By Lemma 3.1 (i) and Fact 2, we have

$$\bigcap_{m \geq n} N_{m,n}(\mathcal{E}_m^\chi) \cong \mathcal{E}^\chi(\mathcal{U}^\chi)^{\omega_n} / (\mathcal{U}^\chi)^{\omega_n} \cong (\Psi(\mathcal{E}^\chi), \omega_n \Psi(\mathcal{U}^\chi)) / \omega_n \Psi(\mathcal{U}^\chi).$$

For sufficiently large $n \geq 0$, $\Psi(\mathcal{E}^\chi) \supseteq \omega_n(f(T))$. Hence the above Λ -module has a finite submodule isomorphic to $(f(T), \Psi(\mathcal{U}^\chi)) / \Psi(\mathcal{U}^\chi)$. However, \mathcal{E}_n^χ is \mathbb{Z}_p -torsion free (Lemma 3.3 (ii)). Thus we have $f(T) \in \Psi(\mathcal{U}^\chi)$.

By Lemma 3.3 (ii), $\mathcal{E}^\chi \cap \mathbb{T}^\chi = \{1\}$. Then $\mathcal{U}^\chi/\mathcal{E}^\chi$ has a Λ -submodule isomorphic to \mathbb{T}^χ , which we also denote by \mathbb{T}^χ . By Fact 2, we have an exact sequence

$$1 \longrightarrow \mathbb{T}^\chi \longrightarrow \mathcal{U}^\chi/\mathcal{E}^\chi \longrightarrow \Lambda/\Psi(\mathcal{E}^\chi) \longrightarrow \Lambda/(T - q, 1 - \chi\omega^{-1}(p)) \longrightarrow 1.$$

Then $(\mathcal{U}^\chi/\mathcal{E}^\chi)/\mathbb{T}^\chi$ has a finite Λ -submodule isomorphic to $(f(T))/\Psi(\mathcal{E}^\chi)$, since we have proved that $f(T) \in (T - q, 1 - \chi\omega^{-1}(p))$. By (3.1), we can regard $(\mathcal{U}^\chi/\mathcal{E}^\chi)/\mathbb{T}^\chi$ as a submodule of \mathfrak{X}/\mathbb{T} . Therefore, by Lemma 3.4, we have $(f(T)) = \Psi(\mathcal{E}^\chi)$. Furthermore, by using the above exact sequence and (3.3), we have $(f(T)) = (\text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi))$. We have proved the lemma. \square

Remark 2. By using Fact 1 (the Iwasawa main conjecture), Fact 2 and the above lemma, we can see that Greenberg's conjecture for (p, χ) is equivalent to the following:

$$\mathcal{E}^\chi = \mathcal{C}^\chi.$$

4. PROOF OF THE MAIN RESULTS

We rewrite the condition $(H_{P,n})$ as follows:

Lemma 4.1. *For each $n \geq 0$, the condition $(H_{P,n})$ is equivalent to the following condition:*

$$(\mathcal{H}_{P,n}) \quad c_n^{\xi_X X_{P,n}} \notin (\mathcal{E}_n^\chi)^{m_{P,n}}.$$

Proof. We fix $n \geq 0$. The isomorphism in Lemma 3.3 maps the class $[c_n^{\mathbf{e}_X \tilde{Y}_{P,n}}]$ in $E'_n/E_n^{m_{P,n}}$ to the class $[c_n^{e_\psi(1-\alpha)\tilde{X}_{P,n}}]$ in $\mathcal{E}_n/\mathcal{E}_n^{m_{P,n}}$. Thus the condition $(H_{P,n})$ holds if and only if

$$c_n^{e_\psi(1-\alpha)\tilde{X}_{P,n}} \notin \mathcal{E}_n^{m_{P,n}}.$$

Putting $\mathfrak{E} = \ker(\sum_{i=0}^{p-1} \alpha^i : \mathcal{E}_n^{e_\psi} \rightarrow \mathcal{E}_n^{e_\psi})$, we have $\mathfrak{E} \cap \mathcal{E}_n^{m_{P,n}} = \mathfrak{E}^{m_{P,n}}$ and $c_n^{e_\psi(1-\alpha)\tilde{X}_{P,n}} \in \mathfrak{E}$. Then the above condition holds if and only if

$$c_n^{e_\psi(1-\alpha)\tilde{X}_{P,n}} \notin \mathfrak{E}^{m_{P,n}}.$$

By using Lemma 3.2, we also see that \mathfrak{E} is isomorphic to \mathcal{E}_n^χ by $x^{e_\psi} \mapsto (x \otimes 1)^{\frac{\xi_X}{1-\alpha}}$ for $x^{e_\psi} \in \mathfrak{E}$. (Note that $\frac{\xi_X}{1-\alpha}$ is in $\mathcal{O}[\Delta]$.) We can see that $\xi_X \tilde{X}_{P,n} = \xi_X X_{P,n}$. Therefore the above condition is equivalent to the condition $(\mathcal{H}_{P,n})$. \square

Proof of Theorem 2.6. We shall show that $P(T) \mid \text{char}_\Lambda A_\infty^\chi$ holds if and only if the opposite

$$(\neg \mathcal{H}_{P,n}) \quad c_n^{\xi_X X_{P,n}} \in (\mathcal{E}_n^\chi)^{m_{P,n}}$$

of $(\mathcal{H}_{P,n})$ holds for all $n \geq 0$.

We put $Q(T) = P_\chi(T)/P(T)$. By Fact 1 (the Iwasawa main conjecture) and (3.2), we have

$$\text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi) \cdot \text{char}_\Lambda A_\infty^\chi = P_\chi(T).$$

Then $P(T) \mid \text{char}_\Lambda A_\infty^\chi$ if and only if $\text{char}_\Lambda(\mathcal{U}^\chi/\mathcal{E}^\chi) \mid Q(T)$. By Lemma 3.5, the latter condition is equivalent to saying that $Q(T) \in \Psi(\mathcal{E}^\chi)$. Let $Q^{(n)}$ denote $Q(T) \bmod \mathfrak{v}_n^\chi$ in $\Lambda/(\mathfrak{v}_n^\chi)$. Here we recall that \mathfrak{v}_n^χ is ω_n (resp. ν_n) if $\chi(p) \neq 1$ (resp. $\chi(p) = 1$). By using the Λ -homomorphism $\Psi_n : \mathcal{V}_n^\chi \rightarrow \Lambda/(\mathfrak{v}_n^\chi)$ defined after Lemma 3.1, we have $Q(T) \in \Psi(\mathcal{E}^\chi)$ if and only if

$$(4.1) \quad Q^{(n)} \in \Psi_n(\mathcal{E}_n^\chi \cap \mathcal{V}_n^\chi)$$

for all $n \geq 0$. For a fixed $n \geq 0$, it suffices to show that (4.1) is equivalent to $(\neg \mathcal{H}_{P,n})$. Since $\Lambda/(\mathfrak{v}_n^\chi)$ is \mathbb{Z}_p -torsion free, the condition (4.1) holds if and only if

$$(4.2) \quad m_{P,n} Q^{(n)} \in \Psi_n((\mathcal{E}_n^\chi \cap \mathcal{V}_n^\chi)^{m_{P,n}}).$$

We defined $X_{P,n}$ to satisfy $m_{P,n} Q(T) \equiv X_{P,n}(T) P_\chi(T) \bmod \mathfrak{v}_n^\chi$. By Fact 2, we have $\Psi_n(c_n^{\xi_X}) = g_\chi(T) = u_\chi(T) P_\chi(T)$ where u_χ is the unit of Λ appearing in (2.1). Hence we have

$$(4.3) \quad u_\chi m_{P,n} Q^{(n)} = \Psi_n(c_n^{\xi_X X_{P,n}}).$$

By Fact 2 and Lemma 3.1 (i), the kernel of Ψ_n is \mathbb{T}_n . Hence, the condition (4.2) holds if and only if

$$(4.4) \quad c_n^{\xi_X X_{P,n}} \in (\mathcal{E}_n^\chi \cap \mathcal{V}_n^\chi)^{m_{P,n}} \mathbb{T}_n.$$

If we assume that the condition (4.4) holds, there exists $\epsilon \in \mathcal{E}_n^\chi \cap \mathcal{V}_n^\chi$ such that $(c_n^{\xi_\chi X_{P,n}}/\epsilon^{m_{P,n}}) \in \mathbb{T}_n$. Since $c_n^{\xi_\chi} \in \mathcal{E}_n^\chi$ and $\mathcal{E}_n^\chi \cap \mathbb{T}_n = \{1\}$ (Lemma 3.3 (ii)), $c_n^{\xi_\chi X_{P,n}} = \epsilon^{m_{P,n}}$. Thus the condition (4.4) holds if and only if

$$(4.5) \quad c_n^{\xi_\chi X_{P,n}} \in (\mathcal{E}_n^\chi)^{m_{P,n}} \cap (\mathcal{V}_n^\chi)^{m_{P,n}}.$$

In the case where $\chi\omega^{-1}(p) \notin \mu_{p^\infty}$, we have $\Psi(\mathcal{U}^\chi) = \Lambda$ by Fact 2. Then, clearly $Q(T) \in \Psi(\mathcal{U}^\chi)$, and hence $Q^{(n)} \in \Psi(\mathcal{V}_n^\chi)$ for all $n \geq 0$. In this case, we have $\mathbb{T}_n = \{1\}$. Thus, by (4.3), we obtain $c_n^{\xi_\chi X_{P,n}} \in (\mathcal{V}_n^\chi)^{m_{P,n}}$. Therefore, the condition (4.5) is equivalent to $(-\mathcal{H}_{P,n})$. If we assume $\chi\omega^{-1}(p) \in \mu_{p^\infty}$, then $\chi(p) \notin \mu_{p^\infty}$. Hence, by Lemma 3.1 (ii), we have $\mathcal{V}_n^\chi = \mathcal{U}_n^\chi$. Thus, $(\mathcal{E}_n^\chi)^{m_{P,n}} \cap (\mathcal{V}_n^\chi)^{m_{P,n}} = (\mathcal{E}_n^\chi)^{m_{P,n}}$. This completes the proof. \square

Proof of Theorem 2.2. The assertion follows from Theorem 2.6 and the following. \square

Lemma 4.2. *Let $P(T)$ be a factor of $P_\chi(T)$. Then we have that $P(T)$ satisfies (T) if and only if $c_n^{\xi_\chi X_{P,n}} \notin (\mathcal{U}_n^\chi)^{m_{P,n}} \mathbb{T}_n$ for some $n \geq 0$.*

Proof. By the definition, $P(T)$ satisfies (T) if and only if $Q(T) \notin (T-q, 1-\chi\omega^{-1}(p))$ with $Q = P_\chi/P$. By Fact 2, we have $\Psi(\mathcal{U}^\chi) = (T-q, 1-\chi_1(p))$. Hence $P(T)$ satisfies (T) if and only if $Q(T) \notin \Psi(\mathcal{U}^\chi)$. Furthermore, by using an argument in the proof of Theorem 2.6, we can show that $Q(T) \notin \Psi(\mathcal{U}^\chi)$ holds if and only if $c_n^{\xi_\chi X_{P,n}} \notin (\mathcal{V}_n^\chi)^{m_{P,n}} \mathbb{T}_n$ for some $n \geq 0$. By Lemmas 2.3 (i) and 3.1 (ii), if $P(T)$ satisfies (T) , then $\mathcal{V}_n^\chi = \mathcal{U}_n^\chi$. Thus, the lemma has been proved. \square

Remark 3. Assume $P(T)$ satisfies (T) . Then we can directly show that $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$. Indeed, we have $Q(T) \notin \Psi(\mathcal{U}^\chi)$ with $Q = P_\chi/P$ by the assumption and Fact 2, and hence $Q(T) \notin \Psi(\mathcal{E}^\chi)$. The latter condition is equivalent to saying that $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$ by Fact 1, Lemma 3.5 and (3.2) (see at the beginning of the proof of Theorem 2.6).

5. BASIC FACTS ON THE χ -PARTS OF THE IDEAL CLASS GROUPS

In this section, we shall show sufficient conditions for A_∞^χ to be trivial (Lemma 5.1 and Remark 1), which are more or less known. We use the same notation as in the previous sections.

For a $\mathbb{Z}_p[\Delta]$ -module M , we define an \mathcal{O} -module M_χ , the χ -quotient, by

$$M_\chi := (M \otimes_{\mathbb{Z}_p} \mathcal{O})/I_\chi(M \otimes_{\mathbb{Z}_p} \mathcal{O}),$$

where I_χ denotes the ideal of $\mathcal{O}[\Delta]$ generated by all elements of the form $\delta - \chi(\delta)$, $\delta \in \Delta$. If M is finite, then M^χ and M_χ have the same orders (cf., e.g., [18, §2]). Thus we have

$$(5.1) \quad \#A_n^\chi = \#A_{n,\chi}.$$

We first show the following:

Lemma 5.1. *Let χ be a (not necessarily even) Dirichlet character with $\chi(p) \notin \mu_{p^\infty}$. If A_0^χ is trivial, then, for all $n \geq 0$, A_n^χ are trivial, and so is A_∞^χ .*

Proof. By Nakayama's lemma and (5.1), it suffices to show that

$$A_{n,\chi} \cong A_{\infty,\chi}/\omega_n A_{\infty,\chi}$$

for each $n \geq 0$. Let L be the maximal unramified abelian p -extension of k_∞ , L_n the maximal abelian extension of k_n contained in L , and H_n the Hilbert p -class field of k_n for each $n \geq 0$. By class field theory, we have $A_n \cong \text{Gal}(H_n/k_n)$ and $A_\infty \cong \text{Gal}(L/k_\infty)$. We further see that $\text{Gal}(L_n/k_\infty) \cong A_\infty/\omega_n A_\infty$, and hence $\text{Gal}(L_n/k_\infty)_\chi \cong A_{\infty,\chi}/\omega_n A_{\infty,\chi}$. Then we show that $\text{Gal}(L_n/k_\infty)_\chi \cong \text{Gal}(H_n/k_n)_\chi$.

For a prime ideal \mathfrak{p} of k_n over p , let $I_{\mathfrak{p}}$ denote the inertia group of \mathfrak{p} in $\text{Gal}(L_n/k_n)$. Since L_n/k_n is unramified outside p , we have an exact sequence of $\mathbb{Z}_p[\text{Gal}(k_n/\mathbb{Q})]$ -modules

$$\prod_{\mathfrak{p}|p} I_{\mathfrak{p}} \longrightarrow \text{Gal}(L_n/k_n) \longrightarrow \text{Gal}(H_n/k_n) \longrightarrow 0.$$

We further see that $\prod_{\mathfrak{p}|p} I_{\mathfrak{p}} \cong \mathbb{Z}_p[\Delta/D]$ (recall that D is the decomposition group of p in Δ). We recall that Δ' is the prime-to- p part of Δ , $\psi = \chi|_{\Delta'}$ and $e_\psi \in \mathbb{Z}_p[\Delta']$ is its idempotent. By the assumption that $\chi(p) \notin \mu_{p^\infty}$, i.e., $\psi(p) \neq 1$, we have $e_\psi(\mathbb{Z}_p[\Delta/D]) = \{1\}$, and hence $e_\psi \text{Gal}(L_n/k_n) \cong e_\psi \text{Gal}(H_n/k_n)$. Therefore we obtain $\text{Gal}(L_n/k_n)_\chi \cong \text{Gal}(H_n/k_n)_\chi$ (see [18, §2]). Since $\text{Gal}(k_\infty/k_n)_\chi = \{1\}$, we have proved the claim. \square

Applying Theorems 2.2 and 2.6 to verify the conjecture for χ , we have only to consider the case where $\chi(p) \in \mu_{p^\infty}$ or A_0^χ is nontrivial by the above lemma. We give a sufficient condition for A_0^χ to be nontrivial.

Lemma 5.2. *Let χ be a Dirichlet character and ψ the prime-to- p order part of χ . If $e_\psi A_{k_\psi}$ is nontrivial, then so is A_0^χ , where A_{k_ψ} is the p -Sylow subgroup of the ideal class group of k_ψ .*

Proof. Assume that $e_\psi A_{k_\psi}$ is nontrivial. Let ρ be the Dirichlet character with $\chi = \psi\rho$. Since k_ρ/\mathbb{Q} is a cyclic extension of degree a power of p , at least one prime l is totally ramified, and hence a prime ideal of k_ψ above l is also totally ramified in $k_\chi = k_\psi k_\rho$. Then we see that the norm map from the p -Sylow subgroup of the ideal class group A_{k_χ} of k_χ to A_{k_ψ} is surjective by class field theory. Therefore, $(e_\psi A_{k_\chi})_{\Delta_p} \rightarrow e_\psi A_{k_\psi}$ is also surjective; so $(e_\psi A_{k_\chi})_{\Delta_p}$ is nontrivial. Let δ_p denote a generator of the p -Sylow subgroup Δ_p of Δ and put $\rho = \chi|_{\Delta_p}$. By the isomorphism

$$A_{0,\chi} = (A_{k_\chi})_\chi \cong (e_\psi A_{k_\chi} \otimes_{\mathcal{O}'} \mathcal{O})/(\delta_p - \rho(\delta_p))(e_\psi A_{k_\chi} \otimes_{\mathcal{O}'} \mathcal{O})$$

(cf. [18, §2]), we have

$$((e_\psi A_{k_\chi})_{\Delta_p} \otimes_{\mathcal{O}'} \mathcal{O})/(1 - \rho(\delta_p))((e_\psi A_{k_\chi})_{\Delta_p} \otimes_{\mathcal{O}'} \mathcal{O}) \cong A_{0,\chi}/(1 - \rho(\delta_p))A_{0,\chi},$$

where \mathcal{O}' denotes the ring generated by the values of ψ over \mathbb{Z}_p . Therefore, $A_{0,\chi}$ is nontrivial, as desired. \square

Finally we shall show that if $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$, then A_n^χ are trivial for all $n \geq 0$ (Remark 1). Let $k_{\omega\chi^{-1}}$ denote the fixed field of the kernel of the odd Dirichlet character $\omega\chi^{-1}$; so $\omega\chi^{-1}$ is a character of $\text{Gal}(k_{\omega\chi^{-1}}/\mathbb{Q})$. The p -Sylow subgroup of the ideal class group of the n -layer $k_{\omega\chi^{-1},n}$ of the cyclotomic \mathbb{Z}_p -extension $k_{\omega\chi^{-1},\infty}/k_{\omega\chi^{-1}}$ is a $\mathbb{Z}_p[\text{Gal}(k_{\omega\chi^{-1}}/\mathbb{Q})]$ -module and we define $A_n^{\omega\chi^{-1}}$ by its $\omega\chi^{-1}$ -part for each $n \geq 0$. As a consequence of the Iwasawa main conjecture proved in [16], the following was proved.

Theorem 5.3 ([16, Theorem 2 in §1.10], [17, Theorem II.1]). *Let χ be a nontrivial even Dirichlet character. Then we have*

$$\#A_0^{\omega\chi^{-1}} = \#(\mathcal{O}/B_{1,\chi\omega^{-1}}\mathcal{O}).$$

We know the following reflection theorem (cf. [20, Theorem 10.9]).

Lemma 5.4. *Let χ be an even Dirichlet character. Then the order of $A_{n,\chi}/\pi A_{n,\chi}$ divides that of $A_n^{\omega\chi^{-1}}/\pi A_n^{\omega\chi^{-1}}$ for each $n \geq 0$, where π is a prime element of \mathcal{O} .*

Assume $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$. By the above and (5.1), we have that $A_0^{\omega\chi^{-1}}$ and A_0^{χ} are trivial. For any character χ , we have $\chi\omega^{-1}(p) \notin \mu_{p^\infty}$ or $\chi(p) \notin \mu_{p^\infty}$. Then, by Lemma 5.1, $A_n^{\omega\chi^{-1}}$ or A_n^{χ} are trivial for all $n \geq 0$. In the former case, the claim follows from Lemma 5.4.

6. EXAMPLES

In this section, we verify Greenberg's conjecture for several examples by using Corollary 2.4 and Theorem 2.6. The computations in this section was carried out by Kazuo Matsuno whom the author wants to thank.

Corollary 2.4 asserts that the condition that $P_\chi(T)$ is irreducible or that $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$ implies $\lambda_\chi = 0$ when $\chi\omega^{-1}(p) \in \mu_{p^\infty} \setminus \{1\}$. We give some examples satisfying the condition.

The above condition is clearly satisfied when $\lambda_\chi^* = \deg P_\chi(T) = 1$. In [19, Lemmas 3 and 4], the author showed that, for each even Dirichlet character ψ of the first kind, there exist infinitely many characters ρ of p -power order such that $\lambda_{\psi\rho}^* = \lambda_\psi^*$ and $\rho(p) \neq 1$. Then $\chi = \psi\rho$ satisfies the condition in Corollary 2.4 if $\lambda_\psi^* = 1$ and $\psi\omega^{-1}(p) = 1$. For example, these conditions are satisfied if $p = 3$ and ψ is the quadratic character of conductor 33 (cf., e.g., [2] for other examples). Therefore there exist infinitely many characters χ satisfying the condition in Corollary 2.4, and hence $\lambda_\chi = 0$. In [18], a method for finding infinitely many χ 's satisfying $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$ and $\chi\omega^{-1}(p) \in \mu_{p^\infty} \setminus \{1\}$ was also given. However, for such characters χ , we can also verify that $\lambda_\chi = 0$ by Lemma 5.1 instead of Corollary 2.4 since we see that $A_0^\chi = 0$ if $\lambda_\chi^* = 1$ or $v_{\mathcal{O}}(B_{1,\chi\omega^{-1}}) = 0$ (for the latter see §5). So we next give examples such that A_0^χ is nontrivial and $P_\chi(T)$ is irreducible.

Let $p = 3$ and ψ be the quadratic character of conductor $321 = 3 \cdot 107$. For a character ρ of order 3, we put $\chi = \psi\rho$. We see that $e_\psi A_{k_\psi} = A_{\mathbb{Q}(\sqrt{321})}$ is nontrivial (the class number of $\mathbb{Q}(\sqrt{321})$ is 3), and hence A_0^χ is also nontrivial by Lemma 5.2. Let ρ be a character of order 3 whose conductor is 7. Then $\rho(3) \neq 1$; so $\chi\omega^{-1}(3) \in \mu_3 \setminus \{1\}$. For this χ , we calculate $P_\chi(T)$ modulo a power of p as in [10] and [5] by using the fact that the Stickelberger elements produce $g_\chi(T)$ and check the irreducibility of $P_\chi(T)$. More precisely, we calculate $NP_\chi(T) = P_\chi(T)P_{\chi^{-1}}(T) \in \mathbb{Z}_3[T]$ (χ^{-1} is a \mathbb{Q}_3 -conjugate of χ), since it is easier to handle \mathbb{Z}_3 -coefficient polynomials than $\mathcal{O} = \mathbb{Z}_3[\zeta_3]$ -coefficient ones. We have

$$NP_\chi(T) \equiv T^4 + 5130T^3 + 1020T^2 + 2214T + 4977 \pmod{3^8}.$$

We see that $NP_\chi(T)$ is irreducible over \mathbb{Z}_3 . So $P_\chi(T)$ is also irreducible over \mathcal{O} . Thus, this χ satisfies the condition in Corollary 2.4, and we have $\lambda_\chi = 0$. Moreover, this implies that Greenberg's conjecture for k_χ , the cyclic field of degree 6 whose conductor is $2247 = 3 \cdot 7 \cdot 107$, is true by Lemma 2.1. In fact, Ichimura and Sumida

[10, Proposition] showed that $\lambda_\psi = 0$, and we can see that $\lambda_\rho = 0$ by Iwasawa's lemma (cf. [20, Lemma 10.4]) since $\rho(3) \neq 1$.

We further check the irreducibility of $P_{\psi\rho}(T)$ for a nontrivial character ρ of cyclic cubic fields k_ρ with conductor less than 10^3 . There are 128 such fields, and 86 fields among them satisfy $\rho(3) \neq 1$ (i.e., 3 does not decompose in those fields). Among these 86 fields, $P_{\psi\rho}(T)$ is irreducible for those ρ whose conductor is in the following list:

$$\begin{aligned} &7, 43, 97, 109, 127, 139, 157, 181, 211, 229, 277, 337, \\ &349, 379, 409, 607, 691, 709, 733, 739, 751, 877, 907, \\ &217 = 7 \cdot 31, 301 = 7 \cdot 43, 427^\dagger = 7 \cdot 61, 469^\dagger = 7 \cdot 67, \\ &511^\dagger = 7 \cdot 73, 553 = 7 \cdot 79, 721^\dagger = 7 \cdot 103. \end{aligned}$$

We remark that there are two cyclic cubic fields k_ρ satisfying $\rho(3) \neq 1$ whose conductor is one of the numbers with \dagger in the above list. (We have only one such field for others.) Among them, $P_{\psi\rho}(T)$ are irreducible for both fields of conductor 427, 511, 721 and for one of the fields of conductor 469. Thus we obtain 33 cyclic cubic fields k_ρ of conductor less than 10^3 such that $\chi = \psi\rho$ satisfies the assumption of Corollary 2.4.

We can see that $\lambda_\rho = 0$ if the conductor of ρ is prime and $\rho(3) \neq 1$ by Iwasawa's lemma. Furthermore, Yamamoto kindly informed us that we know that $\lambda_\rho = 0$ for ρ whose conductor is 217, 301, 427, 469 or 721 (see below for 553). Thus, for such ρ 's, Greenberg's conjecture for $k_\chi = k_\rho(\sqrt{321})$ is true by Lemma 2.1.

We apply Theorem 2.6 to several examples.

Let $p = 3$, k_χ be a unique cubic field of conductor $553 = 7 \cdot 79$ in which 3 does not decompose and χ a nontrivial character of k_χ . We fix a generator σ of $\Delta_p (= \Delta)$ and put $\chi(\sigma) = \zeta$, a nontrivial p -th root of unity. We have

$$P_\chi(T) \equiv T - (122 + 160\zeta) \pmod{3^5}.$$

We check the condition $(H_{P_{\chi,2}})$. We see that $m_{P_{\chi,2}} = 3^3 = 27$ and

$$\begin{aligned} \tilde{Y}_{P_{\chi,2}} &\equiv T^8 + (23 + 25\zeta)T^7 + (3 + 3\zeta)T^6 + (24 + 15\zeta)T^5 \\ &\quad + (6 + 3\zeta)T^4 + 9\zeta T^3 + (21 + 9\zeta)T^2 + (24 + 21\zeta)T + 9 + 18\zeta \pmod{3^3}. \end{aligned}$$

Let $l = 29863$. Then a prime ideal \mathfrak{l} of k_2 above l is of degree one. We verify that $c_2^{(1-\sigma)} \tilde{Y}_{P_{\chi,2}} \pmod{\mathfrak{l}} \notin (\mathbb{Z}/l\mathbb{Z})^{\times 27}$, and so we have $\lambda_\chi = 0$. Greenberg's conjecture for k_χ (and $k_\chi(\sqrt{321})$ also) is also true.

Let $p = 5$ and χ be a character of order 5 with prime conductor f . By a result of Kurihara [15, §4.4], we know that $\lambda_\chi = 0$ for all $f < 10^5$ except $f = 38851, 41201, 84551$. We consider the case $f = 38851$. By computing $NP_\chi(T) = \prod_{i=1}^4 P_{\chi^i}(T) \in \mathbb{Z}_5[T]$, we see that $P_\chi(T)$ decomposes into a product of two irreducible factors $P_1(T)$ and $P_2(T)$ of degrees 1 and 3 respectively. In [15], Kurihara introduced an invariant κ associated to χ and showed that, for a factor $P(T)$ of $P_\chi(T)$, we have $P(T) \nmid \text{char}_\Lambda A_\infty^\chi$ if $\deg P(T) \geq \kappa$ (cf. [15, Remark 1.11]). We have $\kappa = 2$ in this case. So we check the conditions in Theorem 2.6 only for the factor $P_1(T)$ of degree 1. We have

$$P_1(T) \equiv T - (1313 + 416\zeta + 1834\zeta^2 + 2427\zeta^3) \pmod{5^5}.$$

We check the condition $(H_{P_1,1})$. We see that $m_{P_1,1} = 5$ and

$$\tilde{Y}_{P_1,1} = T^3 + (3 + \sigma + 4\sigma^2 + 2\sigma^3)T^2.$$

We verify that $c_1^{(1-\sigma)\tilde{Y}_{P_1,1}} \bmod \mathfrak{l} \notin (\mathbb{Z}/l\mathbb{Z})^{\times 5}$ with $l = 9712751$, where a prime ideal \mathfrak{l} of k_1 above l is of degree one.

We further consider the cases where $f = 41201, 84551$ in the above setting. We see that $P_\chi(T)$ decomposes into a product of two irreducible factors $P_1(T)$ and $P_2(T)$ of degrees 1 and 3 respectively, in both cases, and have $\kappa = 2$ (resp. 3) if $f = 41201$ (resp. 84551). Hence we have only to check the conditions in Theorem 2.6 for the factor $P_1(T)$ of degree 1 for both f 's. We verify that the condition $(H_{P_1,1})$ holds in each case. Therefore, together with the result of Kurihara [15, §4.4], we have that $\lambda_\chi = 0$ for $p = 5$ and all primes $f < 10^5$.

ACKNOWLEDGEMENTS

The author would like to express her sincere gratitude to Professor Humio Ichimura and Doctor Hiroki Sumida for various advice and for encouragement. She also would like to thank Doctor Kazuo Matsuno for carrying out calculations in §6 which allowed us to verify Greenberg's conjecture for several examples, and thank Doctor Gen Yamamoto for kindly informing us about known results on the conjecture.

REFERENCES

- [1] A. Brumer, *On the units of algebraic number fields*, *Mathematika* **14** (1967) 121–124. MR **36**:3746
- [2] D. S. Dummit, D. Ford, H. Kisilevsky and W. Sands, *Computation of Iwasawa lambda invariants for imaginary quadratic fields*, *J. Number Theory* **37** (1991) 100–121. MR **92a**:11124
- [3] B. Ferrero and R. Greenberg, *On the behavior of p -adic L -functions at $s = 0$* , *Invent. Math.* **50** (1978) 91–102. MR **80f**:12016
- [4] B. Ferrero and L. C. Washington, *The Iwasawa invariant μ_p vanishes for abelian number fields*, *Ann. of Math. (2)* **109** (1979) 377–395. MR **81a**:12005
- [5] T. Fukuda and K. Komatsu, *Ichimura-Sumida criterion for Iwasawa λ -invariants*, *Proc. Japan Acad. Ser. A Math. Sci.* **76** (2000) 111–115. MR **2001g**:11168
- [6] R. Gillard, *Unités cyclotomiques, unités semi-locales et \mathbb{Z}_ℓ -extensions II*, *Ann. Inst. Fourier (Grenoble)* **29** (1979) 49–79. MR **81e**:12005a
- [7] R. Greenberg, *On the Iwasawa invariants of totally real number fields*, *Amer. J. Math.* **98** (1976) 263–284. MR **53**:5529
- [8] H. Ichimura and H. Sumida, *On the Iwasawa invariants of certain real abelian fields*, *Tôhoku Math. J.* **49** (1997) 203–215. MR **98e**:11128a
- [9] H. Ichimura and H. Sumida, *On the Iwasawa λ -invariant of the real p -cyclotomic field*, *J. Math. Sci. Univ. Tokyo* **3** (1996) 457–470. MR **98e**:11128b
- [10] H. Ichimura and H. Sumida, *On the Iwasawa invariants of certain real abelian fields II*, *Internat. J. Math.* **7** (1996) 721–744. MR **98e**:11128c
- [11] K. Iwasawa, *On some modules in the theory of cyclotomic fields*, *J. Math. Soc. Japan* **16** (1964) 42–82. MR **35**:6646
- [12] K. Iwasawa, *On \mathbb{Z}_ℓ -extensions of algebraic number fields*, *Ann. of Math. (2)* **98** (1973) 246–326. MR **50**:2120
- [13] J. Kraft and R. Schoof, *Computing Iwasawa modules of real quadratic number fields*, *Compositio Math.* **97** (1995) 135–155. MR **97b**:11129
- [14] M. Kurihara, *The Iwasawa λ -invariants of real abelian fields and the cyclotomic elements*, *Tokyo J. Math.* **22** (1999) 259–277. MR **2001a**:11182
- [15] M. Kurihara, *Remarks on the λ_p -invariants of cyclic fields of degree p* , preprint.
- [16] B. Mazur and A. Wiles, *Class fields of abelian extensions of \mathbb{Q}* , *Invent. Math.* **76** (1984) 179–330. MR **85m**:11069

- [17] D. Solomon, *On the classgroups of imaginary abelian fields*, Ann. Inst. Fourier (Grenoble) **40** (1990) 467–492. MR **92a**:11133
- [18] T. Tsuji, *Semi-local units modulo cyclotomic units*, J. Number Theory **78** (1999) 1–26. MR **2000f**:11148
- [19] T. Tsuji, *Greenberg's conjecture for Dirichlet characters of order divisible by p* , Proc. Japan Acad. Ser. A Math. Sci. **77** (2001) 52–54. MR **2002d**:11130
- [20] L. C. Washington, *Introduction to Cyclotomic Fields*, Graduate Texts in Math. 83, Springer-Verlag, New York, 1982. MR **85g**:11001

DEPARTMENT OF MATHEMATICS, TOKAI UNIVERSITY, HIRATSUKA, KANAGAWA, 259-1292, JAPAN
E-mail address: tsuji@sm.u-tokai.ac.jp

PSEUDO-HOLOMORPHIC CURVES IN COMPLEX GRASSMANN MANIFOLDS

XIAOXIANG JIAO AND JIAGUI PENG

ABSTRACT. It is proved that the Kähler angle of the pseudo-holomorphic sphere of constant curvature in complex Grassmannians is constant. At the same time we also prove several pinching theorems for the curvature and the Kähler angle of the pseudo-holomorphic spheres in complex Grassmannians with non-degenerate associated harmonic sequence.

1. INTRODUCTION

In this paper we study conformal minimal two-spheres in complex Grassmann manifolds by using the harmonic sequence. Given a harmonic map φ of surfaces M into the complex Grassmannian $\mathbf{G}_{k,n}$, by using the ∂' -transform Chern and Wolfson ([3], [10]) obtained the following harmonic sequence associated to φ :

$$\varphi = \varphi_0 \xrightarrow{\partial'} \varphi_1 \xrightarrow{\partial'} \cdots \xrightarrow{\partial'} \varphi_j \xrightarrow{\partial'} \cdots,$$

where $\varphi_{j+1} = \partial' \varphi_j$, $j = 0, 1, \dots$, and $\varphi_j : M \rightarrow \mathbf{G}_{k_j,n}$ are harmonic maps, $k_j = \text{rank}(\varphi_j)$. If φ_j is anti-holomorphic, then $k_{j+1} = 0$. When φ is holomorphic we call φ_j a pseudo-holomorphic curve generated by φ . Such curves with the induced metrics from the associated complex Grassmann manifolds form a class of minimal immersions. When $k_j = k_{j+1}$ we say that φ_j is *non-degenerate*. When $k_j = k_{j+1}$ for all j we say that the harmonic sequence associated to the map φ is *non-degenerate*.

When specialized to $\mathbf{G}_{1,n} = \mathbf{CP}^{n-1}$, any pseudo-holomorphic curve is obtained from a holomorphic curve projected into \mathbf{CP}^{n-1} . Calabi ([2]) showed that any simply connected holomorphic curve in \mathbf{CP}^{n-1} is completely determined, up to holomorphic isometries of \mathbf{CP}^{n-1} , by its induced metric. Calabi also showed that a simply connected holomorphic curve of constant curvature in \mathbf{CP}^{n-1} is the Veronese curve, up to unitary equivalence. For a pseudo-holomorphic curve in \mathbf{CP}^{n-1} , Bolton, Jensen, Rigoli and Woodward ([1]) showed that, up to a holomorphic isometry of \mathbf{CP}^{n-1} , the harmonic sequence determined by any linearly full conformal minimal immersion of constant curvature in \mathbf{CP}^{n-1} is the Veronese sequence, in which each map is a minimal immersion with constant curvature and constant Kähler angle.

Received by the editors September 6, 2002 and, in revised form, October 31, 2002.

2000 *Mathematics Subject Classification*. Primary 53C42, 53C55.

Key words and phrases. Gauss curvature, Kähler angle, harmonic sequence, pseudo-holomorphic curve.

Supported by the National Natural Science Foundation of China (Grants No. 10001033, 10131020, 10071804) and the President Foundation of the Graduate School of the Chinese Academy of Sciences.

It is well known that the rigidity fails for pseudo-holomorphic curves or holomorphic curves generalized to $\mathbf{G}_{k,n}$ ([5], [14]). For example, Chi and Zheng ([5]) classified the holomorphic curves of the Riemann sphere into $\mathbf{G}_{2,4}$ with the induced constant curvature 2 into two classes, up to unitary equivalence, in which none of the curves are congruent. Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a pseudo-holomorphic curve in a complex Grassmannian $\mathbf{G}_{k,n}$. Problem: *Is the Kähler angle $\theta(\varphi)$ of φ constant when its Gauss curvature $K(\varphi)$ is constant? What are the relationships between the Kähler angle and the Gauss curvature of φ and its ramification index?* In this paper we will investigate these questions.

In the second and third sections of this paper we obtain some fundamental formulas for pseudo-holomorphic curves in complex Grassmann manifolds.

In the fourth section, by using these formulas we prove that the curvatures of pseudo-holomorphic curves are equal to $4/N$ (N is a positive integer) if these curvatures are constant (this result was proved by Chi and Zheng in [5]) (Theorem 4.1), and prove that Kähler angles of pseudo-holomorphic curves of constant curvature are constant (Theorem 4.2). In this section, we also give a harmonic sequence, in which each map is a minimal immersion with constant curvature and constant Kähler angle.

In the final section, we give some pinching theorems for pseudo-holomorphic curves with the associated non-degenerate harmonic sequence for curvatures and Kähler angles (Theorems 5.2, 5.6 and 5.7). At the same time we also show that the Kähler angle of a pseudo-holomorphic curve is independent of its ramification index under the assumption of Theorem 5.2.

2. MINIMAL IMMERSIONS AND HARMONIC SEQUENCES

Let $U(n)$ be the unitary group. Let M be a simply connected domain in the unit sphere \mathbf{S}^2 and let (z, \bar{z}) be a complex coordinate on M . We take the metric $ds_M^2 = dzd\bar{z}$ on M . Denote

$$\partial = \frac{\partial}{\partial z}, \quad \bar{\partial} = \frac{\partial}{\partial \bar{z}}, \quad A_z = \frac{1}{2}s^{-1}\partial s, \quad A_{\bar{z}} = \frac{1}{2}s^{-1}\bar{\partial} s.$$

Let $s : M \rightarrow U(n)$ be a smooth map; then s is a harmonic map if and only if it satisfies the following equation ([9]):

$$(1) \quad \bar{\partial} A_z = [A_z, A_{\bar{z}}].$$

If $s : \mathbf{S}^2 \rightarrow U(n)$ is a harmonic map, then s is a conformal map; so s is a minimal immersion. Let $\omega = g^{-1}dg$ be a Maurer-Cartan form on $U(n)$, and let $ds_{U(n)}^2 = \frac{1}{8} \operatorname{tr} \omega \omega^*$ be the metric on $U(n)$. Then the metric induced by s on \mathbf{S}^2 is given by

$$(2) \quad ds^2 = -\operatorname{tr} A_z A_{\bar{z}} dzd\bar{z}.$$

Let $\mathbf{G}_{k,n}$ be the complex Grassmann manifold consisting of all complex k -dimensional subspaces in \mathbf{C}^n . Here we consider $\mathbf{G}_{k,n}$ as the set of Hermitian orthogonal projections onto a k -dimensional subspace in \mathbf{C}^n , i.e., $\mathbf{G}_{k,n} = \{\varphi \text{ is the Hermitian orthogonal projection onto a } k\text{-dimensional subspace in } \mathbf{C}^n\}$. Then $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ is a Hermitian orthogonal projection onto a k -dimensional subbundle $\eta \subset \mathbf{S}^2 \times \mathbf{C}^n$, and $s = \varphi - \varphi^\perp$ is a map from \mathbf{S}^2 into $U(n)$. It is well known that φ is harmonic if and only if s is harmonic. If $\varphi^\perp \bar{\partial} \varphi = 0$ or $\varphi^\perp \partial \varphi = 0$, we call φ a *holomorphic curve* or an *anti-holomorphic curve* in $\mathbf{G}_{k,n}$.

Using φ , the harmonic sequences (see [3], [10]) are given by

$$(3) \quad \varphi = \varphi_0 \xrightarrow{\partial'} \varphi_1 \xrightarrow{\partial'} \cdots \xrightarrow{\partial'} \varphi_\alpha \xrightarrow{\partial'} \cdots,$$

$$(4) \quad \varphi = \varphi_0 \xrightarrow{\partial''} \varphi_{-1} \xrightarrow{\partial''} \cdots \xrightarrow{\partial''} \varphi_{-\alpha} \xrightarrow{\partial''} \cdots,$$

where $\varphi_\alpha : \mathbf{S}^2 \times \mathbf{C}^n \rightarrow \text{Im}(\varphi_{\alpha-1}^\perp \partial \varphi_{\alpha-1})$ and $\varphi_{-\alpha} : \mathbf{S}^2 \times \mathbf{C}^n \rightarrow \text{Im}(\varphi_{-\alpha+1}^\perp \bar{\partial} \varphi_{-\alpha+1})$ are Hermitian orthogonal projections, $\alpha = 1, 2, \dots$.

Proposition 2.1 ([7]). *For (3) and (4), we have*

$$\varphi_\alpha \partial \varphi_\alpha = -\varphi_{\alpha-1}^\perp \partial \varphi_{\alpha-1}, \quad \varphi_\alpha^\perp \bar{\partial} \varphi_\alpha = -\varphi_{\alpha-1} \bar{\partial} \varphi_{\alpha-1},$$

where $\alpha = \pm 1, \pm 2, \dots$.

If φ_0 is a holomorphic curve in (3) or an anti-holomorphic curve in (4), then elements in (3) or (4) are finite and are mutually orthogonal. If there exists a holomorphic curve φ_0 in $\mathbf{G}_{k,n}$ such that φ is an element in the harmonic sequence (3), i.e., $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha,n}$ belongs to the harmonic sequence

$$(5) \quad 0 \xrightarrow{\partial'} \varphi_0 \xrightarrow{\partial'} \varphi_1 \xrightarrow{\partial'} \cdots \xrightarrow{\partial'} \varphi = \varphi_\alpha \xrightarrow{\partial'} \cdots \xrightarrow{\partial'} \varphi_{\alpha_0} \xrightarrow{\partial'} 0,$$

then we call φ a *pseudo-holomorphic curve* in complex Grassmann manifolds, and α_0 is called the *length* of the harmonic sequence (5).

Now we assume that $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha,n}$ is a pseudo-holomorphic curve. Then we may choose the local unitary frame e_1, e_2, \dots, e_n on $\mathbf{S}^2 \times \mathbf{C}^n$ such that $e_{k_{\alpha-1}+1}, \dots, e_{k_\alpha}$ span $\text{Im}(\varphi_{\alpha-1}^\perp \partial \varphi_{\alpha-1})$, where $k_\alpha = \text{rank}(\varphi_{\alpha-1}^\perp \partial \varphi_{\alpha-1})$, $\alpha = 1, 2, \dots$, $k_0 = \text{rank}(\varphi_0)$.

Let $W_\alpha = (e_{k_{\alpha-1}+1}, e_{k_\alpha}, \dots, e_{k_\alpha})$ be an $(n \times k_\alpha)$ -matrix. Then we have

$$(6) \quad \varphi_\alpha = W_\alpha W_\alpha^*,$$

$$(7) \quad W_\alpha^* W_\alpha = I_{k_\alpha \times k_\alpha}, \quad W_\alpha^* W_{\alpha+1} = 0, \quad W_\alpha^* W_{\alpha-1} = 0.$$

By (7) and a straightforward computation we obtain

$$(8) \quad \begin{cases} \partial W_\alpha = W_{\alpha+1} \Omega_\alpha + W_\alpha \Psi_\alpha, \\ \bar{\partial} W_\alpha = -W_{\alpha-1} \Omega_{\alpha-1}^* - W_\alpha \Psi_\alpha^*, \end{cases}$$

where Ω_α is a $(k_{\alpha+1} \times k_\alpha)$ -matrix and Ψ_α is a $(k_\alpha \times k_\alpha)$ -matrix, $\alpha = 0, 1, 2, \dots$.

It is well known that $\Omega_\alpha = 0$ or $\Omega_{\alpha-1} = 0$ in (8) if and only if φ_α is anti-holomorphic or holomorphic. It is very evident that integrability conditions for (8) are

$$(9) \quad \bar{\partial} \Omega_\alpha = \Psi_{\alpha+1}^* \Omega_\alpha - \Omega_\alpha \Psi_\alpha^*,$$

$$(10) \quad \bar{\partial} \Psi_\alpha + \partial \Psi_\alpha^* = \Omega_\alpha^* \Omega_\alpha + \Psi_\alpha^* \Psi_\alpha - \Omega_{\alpha-1} \Omega_{\alpha-1}^* - \Psi_\alpha \Psi_\alpha^*.$$

By (8), $A_z^{(\alpha)}$ and $A_{\bar{z}}^{(\alpha)}$ for φ_α are given by

$$(11) \quad A_z^{(\alpha)} = -W_\alpha \Omega_{\alpha-1} W_{\alpha-1}^* - W_{\alpha+1} \Omega_\alpha W_\alpha^*,$$

$$(12) \quad A_{\bar{z}}^{(\alpha)} = W_\alpha \Omega_\alpha^* W_{\alpha+1}^* + W_{\alpha-1} \Omega_{\alpha-1}^* W_\alpha^*.$$

It can easily be checked that (9) is equivalent to (1). An immediate consequence of (8) is

Proposition 2.2. *Let $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ be a pseudo-holomorphic curve, with Ω_α and Ψ_α determined by equations (8). Then Ω_α and Ψ_α satisfy equations (9) and (10).*

Let $\varphi^{(\alpha)} = \varphi_0 \oplus \cdots \oplus \varphi_\alpha$ for (5) and $k_{(\alpha)} = k_0 + \cdots + k_\alpha$. Then by Proposition 2.1 we have

$$(13) \quad \partial\varphi^{(\alpha)} = \varphi_\alpha^\perp \partial\varphi_\alpha.$$

Hence $\varphi^{(\alpha)} : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_{(\alpha)}, n}$ is a holomorphic map, and the harmonic map sequence (5) becomes

$$(14) \quad 0 \xrightarrow{\partial'} \varphi^{(\alpha)} \xrightarrow{\partial'} \varphi_{\alpha+1} \xrightarrow{\partial'} \cdots \xrightarrow{\partial'} \varphi_{\alpha_0} \xrightarrow{\partial'} 0.$$

If $k_\alpha = k_{\alpha+1}$, i.e., $\text{rank}(\varphi_\alpha) = \text{rank}(\varphi_{\alpha+1})$, then φ_α is called *non-degenerate*. If φ_α is non-degenerate for $\alpha = 0, 1, \dots, \alpha_0 - 1$ in (5), i.e., $k_0 = k_1 = \cdots = k_{\alpha_0}$, then the harmonic sequence (5) is called the *non-degenerate harmonic sequence* associated to the harmonic map $\varphi = \varphi_\alpha$. Now we assume that φ_α is non-degenerate; then $\det(\Omega_\alpha)$ is a well-defined invariant on \mathbf{S}^2 and has only isolated zeros. Let

$$(15) \quad l_\alpha = \text{tr}(\Omega_\alpha \Omega_\alpha^*).$$

Then

$$l_\alpha = \text{tr}(\varphi_\alpha^\perp \partial\varphi_\alpha \bar{\partial}\varphi_\alpha) = \text{tr}(\partial\varphi^{(\alpha)} \bar{\partial}\varphi^{(\alpha)}), \quad l_{\alpha-1} + l_\alpha = -\text{tr}(A_z^{(\alpha)} A_{\bar{z}}^{(\alpha)}),$$

and we have

Proposition 2.3. *If $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is a non-degenerate pseudo-holomorphic curve, then*

$$(16) \quad 2\partial\bar{\partial} \log |\det(\Omega_\alpha)| = l_{\alpha-1} - 2l_\alpha + l_{\alpha+1}.$$

Proof. By (9) and the rule of differentiating a determinant, we get

$$\bar{\partial} \log \det(\Omega_\alpha) = \text{tr}(\Omega_\alpha^{-1} \bar{\partial}\Omega_\alpha) = \text{tr} \Psi_{\alpha+1}^* - \text{tr} \Psi_\alpha^*,$$

$$\partial \log \det(\Omega_\alpha^*) = \text{tr}((\Omega_\alpha^*)^{-1} \partial\Omega_\alpha^*) = \text{tr} \Psi_{\alpha+1} - \text{tr} \Psi_\alpha.$$

It is not difficult to obtain (16) by (10). \square

Remark. If φ_α is non-degenerate for all α in (5), then

$$(17) \quad 2\partial\bar{\partial} \log |\det(\Omega_\alpha)| = l_{\alpha-1} - 2l_\alpha + l_{\alpha+1}$$

for $\alpha = 0, 1, \dots, \alpha_0 - 1$, where $l_{-1} = l_{\alpha_0} = 0$. When $k_\alpha = 1$ for all α , then $l_\alpha = |\det(\Omega_\alpha)|^2$, and (17) is just the *unintegrated Plücker formulae* for l_α derived by Bolton, Jensen, Rigoli and Woodward in [1].

3. KÄHLER ANGLES AND GAUSS CURVATURES

If $\varphi : M \rightarrow \mathbf{G}_{k, n}$ is a conformal immersion of a Riemann surface M , we define the Kähler angle of φ to be the function $\theta : M \rightarrow [0, \pi]$ given in terms of a complex coordinate z on M by

$$(18) \quad \tan \frac{\theta(p)}{2} = \frac{|d\varphi(\partial/\partial\bar{z})|}{|d\varphi(\partial/\partial z)|}, \quad p \in M.$$

It is clear that θ is globally defined and is smooth at p unless $\theta(p) = 0$ or π . Let $z = x + \sqrt{-1}y$, and let J denote the complex structure on $\mathbf{G}_{k, n}$; then θ is the angle between $Jd\varphi(\partial/\partial x)$ and $d\varphi(\partial/\partial y)$. The importance of the Kähler angle in

the theory of minimal immersions of surfaces into Kähler manifolds was pointed out by Chern and Wolfson [4]. Indeed, φ is holomorphic if and only if $\theta(p) = 0$ for all $p \in M$, while φ is anti-holomorphic if and only if $\theta(p) = \pi$ for all $p \in M$.

Now suppose that $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ is a conformal minimal immersion in the harmonic sequence (5). Then each $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha,n}$ is a conformal minimal immersion. So there exists a finite set X_α (see [1]) such that the Kähler angle

$$\theta_\alpha : \mathbf{S}^2 \setminus X_\alpha \rightarrow [0, \pi]$$

is well defined, and is smooth on $\mathbf{S}^2 \setminus X_\alpha$.

Let $t_\alpha = \left(\tan \frac{\theta_\alpha}{2}\right)^2$. Then, in terms of a local complex coordinate z ,

$$(19) \quad t_\alpha = \frac{|d\varphi_\alpha(\partial/\partial\bar{z})|^2}{|d\varphi_\alpha(\partial/\partial z)|^2} = \frac{l_{\alpha-1}}{l_\alpha}.$$

Let ds_α^2 and $ds_{(\alpha)}^2$ be the metrics on $\mathbf{S}^2 \setminus X_\alpha$ induced by φ_α and $\varphi^{(\alpha)}$ respectively. Then by (11), (12) and (13) we have

$$(20) \quad ds_\alpha^2 = (l_{\alpha-1} + l_\alpha)dzd\bar{z}, \quad ds_{(\alpha)}^2 = l_\alpha dzd\bar{z}.$$

The Laplacians Δ_α and $\Delta_{(\alpha)}$ for ds_α^2 and $ds_{(\alpha)}^2$ are given by

$$(21) \quad \Delta_\alpha = \frac{4}{l_{\alpha-1} + l_\alpha} \partial\bar{\partial}, \quad \Delta_{(\alpha)} = \frac{4}{l_\alpha} \partial\bar{\partial},$$

and the curvatures K_α , $K_{(\alpha)}$ of φ_α and $\varphi^{(\alpha)}$ by

$$(22) \quad K_\alpha = -\frac{2}{l_{\alpha-1} + l_\alpha} \partial\bar{\partial} \log(l_{\alpha-1} + l_\alpha), \quad K_{(\alpha)} = -\frac{2}{l_\alpha} \partial\bar{\partial} \log l_\alpha,$$

the area forms dv_α and $dv_{(\alpha)}$ by

$$(23) \quad dv_\alpha = (l_{\alpha-1} + l_\alpha) \frac{d\bar{z} \wedge dz}{2\sqrt{-1}}, \quad dv_{(\alpha)} = l_\alpha \frac{d\bar{z} \wedge dz}{2\sqrt{-1}}.$$

Choose holomorphic sections $f_1, \dots, f_{k_{(\alpha)}}$ in $\Gamma(\mathbf{S}^2 \times \mathbf{C}^n)$ so that they span $\text{Im}(\varphi^{(\alpha)})$ and

$$(23) \quad f_1 \wedge \dots \wedge f_{k_{(\alpha)}} : \mathbf{S}^2 \rightarrow \mathbf{C}^{\binom{n}{k_{(\alpha)}}}$$

is a nowhere zero holomorphic curve.

Let $F^{(\alpha)} = f_1 \wedge \dots \wedge f_{k_{(\alpha)}}$. Now consider the Plücker embedding (see [12], [13])

$$(24) \quad [F^{(\alpha)}] : \mathbf{S}^2 \rightarrow \mathbf{CP}^{\binom{n}{k_{(\alpha)}}-1},$$

which is a holomorphic isometry, and

$$(25) \quad [F^{(\alpha)}]^* ds_{(\alpha)}^2 = l_\alpha dzd\bar{z}.$$

By [1], we have

$$(26) \quad \partial\bar{\partial} \log |F^{(\alpha)}|^2 = l_\alpha,$$

and the degree δ_α of $F^{(\alpha)}$ is given by

$$(27) \quad \delta_\alpha = \frac{1}{2\pi\sqrt{-1}} \int_{\mathbf{S}^2} \partial\bar{\partial} \log |F^{(\alpha)}|^2 d\bar{z} \wedge dz = \frac{1}{2\pi\sqrt{-1}} \int_{\mathbf{S}^2} l_\alpha d\bar{z} \wedge dz,$$

which is equal to the degree of the polynomial function $F^{(\alpha)}$ in z . We call δ_α the degree of the holomorphic curve $\varphi^{(\alpha)}$. Thus from (17) and (27) we get

Proposition 3.1. *If $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is a non-degenerate pseudo-holomorphic curve, then*

$$(28) \quad -\sharp_\alpha = \delta_{\alpha-1} - 2\delta_\alpha + \delta_{\alpha+1},$$

where $\sharp_\alpha = -\frac{1}{\pi\sqrt{-1}} \int_{\mathbf{S}^2} \partial\bar{\partial} \log |\det \Omega_\alpha| d\bar{z} \wedge dz$ is the number of singular points of Ω_α , i.e., the number of zeros of $\det \Omega_\alpha$.

Remark. If φ_α is non-degenerate for $\alpha = 0, 1, \dots, \alpha_0 - 1$, then $-\sharp_\alpha = \delta_{\alpha-1} - 2\delta_\alpha + \delta_{\alpha+1}$ for all α , and $\delta_{-1} = \delta_{\alpha_0} = 0$; in particular, when $k_0 = \dots = k_{\alpha_0} = 1$, (28) is the global Plücker formula (see [6]).

Let $ds^2 = |\det \Omega_\alpha|^2 dz d\bar{z} = \psi_\alpha \bar{\psi}_\alpha$, where ψ_α is a type $(1, 0)$ analytic 1-form. Then $ds^2 = \psi_\alpha \oplus \bar{\psi}_\alpha$ is a singular Hermitian metric. Let $D_S = \sum_{p \in \mathbf{S}^2} \text{ord}_p(\psi_\alpha) p$ be

the singular divisor of (\mathbf{S}^2, ds^2) , i.e., the zero divisor of ψ_α . By the Gauss-Bonnet-Chern theorem we have

$$\sharp_\alpha = \tau_\alpha + 2,$$

where $\tau_\alpha = \deg D_S$.

We say that τ_α is the *ramification index* of φ_α . Evidently, τ_α is a non-negative integer. If $\tau_\alpha = 0$, φ_α is called *unramified* by Bolton et al. ([1]).

Let (5) be the non-degenerate harmonic sequence; if $\tau_\alpha = 0$ for $\alpha = 0, 1, \dots, \alpha_0 - 1$, the harmonic sequence (5) is called *totally unramified*. Let $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ be the pseudo-holomorphic conformal immersion with the non-degenerate associated harmonic sequence (5); we say that φ is a *totally unramified pseudo-holomorphic conformal immersion* if $\varphi_0, \dots, \varphi_{\alpha_0}$ is totally unramified.

If $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is a conformal minimal immersion with constant Kähler angle, then we have

$$(29) \quad t_\alpha = \frac{\delta_{\alpha-1}}{\delta_\alpha},$$

and from (19) and (22) it follows that

$$(30) \quad K_\alpha = -\frac{2}{l_{\alpha-1} + l_\alpha} \partial\bar{\partial} \log l_{\alpha-1} = -\frac{2}{l_{\alpha-1} + l_\alpha} \partial\bar{\partial} \log l_\alpha.$$

4. CONFORMAL MINIMAL IMMERSIONS WITH CONSTANT CURVATURES

It is well known that any complex submanifold of a (simply-connected, complete) space of constant holomorphic curvature is completely determined, up to holomorphic isometries of the ambient space, by its induced metric (see [2], [8]). The Veronese sequence is the harmonic sequence

$$(31) \quad 0 \xrightarrow{\partial'} \varphi_0 \xrightarrow{\partial'} \dots \xrightarrow{\partial'} \varphi_n \xrightarrow{\partial'} 0,$$

where $n = \deg(\varphi_0)$, and each $\varphi_\alpha = [g_{\alpha,0}, \dots, g_{\alpha,n}] : \mathbf{S}^2 \rightarrow \mathbf{CP}^n$ is given by

$$g_{\alpha,j} = \frac{\alpha!}{(1+z\bar{z})^\alpha} \sqrt{\binom{n}{j}} z^{j-\alpha} \sum_k (-1)^k \binom{j}{\alpha-k} \binom{n-j}{k} (z\bar{z})^k, \quad \alpha, j = 0, 1, \dots, n.$$

Each map φ_α in the Veronese sequence (31) is a conformal minimal immersion with constant curvature

$$(32) \quad K(\varphi_\alpha) = \frac{4}{n + 2\alpha(n - \alpha)}$$

and constant Kähler angle θ_α given by

$$(33) \quad \left(\tan \frac{\theta_\alpha}{2} \right)^2 = \frac{\alpha(n - \alpha + 1)}{(\alpha + 1)(n - \alpha)}.$$

Bolton, Jensen, Rigoli and Woodward ([1]) showed that, up to a holomorphic isometry of \mathbf{CP}^n , the harmonic sequence determined by $\varphi : \mathbf{S}^2 \rightarrow \mathbf{CP}^n$, which is a linearly full conformal minimal immersion of constant curvature, is the Veronese sequence. It is very complicated for pseudo-holomorphic curves in complex Grassmann manifolds; for example, rigidity fails, but we still believe that there are some good geometric properties. In this section we discuss pseudo-holomorphic curves of constant curvature in complex Grassmann manifolds, and Kähler angles.

Let $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ be a pseudo-holomorphic curve with constant curvature. Then we know that

$$[F^{(\alpha-1)}] : \mathbf{S}^2 \rightarrow \mathbf{CP}^{(k_{(\alpha-1)})}{}^{-1}, \quad [F^{(\alpha)}] : \mathbf{S}^2 \rightarrow \mathbf{CP}^{(k_{(\alpha)})}{}^{-1}$$

are two holomorphic curves with degrees $\delta_{\alpha-1}$ and δ_α respectively. Consider the tensor product of $[F^{(\alpha-1)}]$ and $[F^{(\alpha)}]$,

$$(34) \quad T^{(\alpha)} = F^{(\alpha-1)} \otimes F^{(\alpha)}.$$

Then

$$[T^{(\alpha)}] : \mathbf{S}^2 \rightarrow \mathbf{CP}^{(k_{(\alpha-1)})}{}^{-1} (k_{(\alpha-1)})^{-1}$$

is a well-defined holomorphic curve, and from (25) the metric induced by $[T^{(\alpha)}]$ is given by

$$[T^{(\alpha)}]^* ds^2_{\mathbf{CP}^{(k_{(\alpha-1)})}{}^{-1} (k_{(\alpha-1)})^{-1}} = [F^{(\alpha-1)}]^* ds^2_{\mathbf{CP}^{(k_{(\alpha-1)})}{}^{-1}} + [F^{(\alpha)}]^* ds^2_{\mathbf{CP}^{(k_{(\alpha)})}{}^{-1}},$$

i.e.,

$$(35) \quad [T^{(\alpha)}]^* ds^2_{\mathbf{CP}^{(k_{(\alpha-1)})}{}^{-1} (k_{(\alpha-1)})^{-1}} = (l_{\alpha-1} + l_\alpha) dz d\bar{z}.$$

Hence the curvature K_α of φ_α is equal to the curvature of $[T^{(\alpha)}]$. From [1], an immediate consequence is

Theorem 4.1. *If $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k, n}$ is a pseudo-holomorphic curve with constant curvature $K(\varphi)$, then $K(\varphi) = 4/N$, where N is a positive integer.*

This theorem was proved by Chi and Zheng ([5]) by the method of the moving frame. In the following we will prove

Theorem 4.2. *If $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is a pseudo-holomorphic curve with constant curvature K_α , then the Kähler angle θ_α of φ_α is constant.*

Proof. From (22) we have

$$(36) \quad K_\alpha(l_{\alpha-1} + l_\alpha) = -2\partial\bar{\partial} \log(l_{\alpha-1} + l_\alpha).$$

When K_α is constant, from (22), (23), (27) and the Gauss-Bonnet theorem it follows that

(37)
$$K_\alpha = \frac{4}{\delta_{\alpha-1} + \delta_\alpha}.$$

Hence from (22) we obtain

(38)
$$-\frac{2}{l_{\alpha-1} + l_\alpha} \partial \bar{\partial} \log(l_{\alpha-1} + l_\alpha) = \frac{4}{\delta_{\alpha-1} + \delta_\alpha}.$$

Choose a complex coordinate z on $\mathbf{S}^2 \setminus \{pt\}$ so that

(39)
$$l_{\alpha-1} + l_\alpha = \frac{\delta_{\alpha-1} + \delta_\alpha}{(1 + z\bar{z})^2}.$$

From (38), (39) and (26) we obtain

(40)
$$\partial \bar{\partial} \log \frac{|F^{(\alpha-1)}|^2 |F^{(\alpha)}|^2}{(1 + z\bar{z})^{\delta_{\alpha-1} + \delta_\alpha}} = 0.$$

Since we can choose holomorphic sections $f_1, \dots, f_{k_{(\alpha)}}$ in $\Gamma(\mathbf{S}^2 \times \mathbf{C}^n)$ such that the maps $F^{(\alpha-1)}$ and $F^{(\alpha)}$ are polynomial functions on \mathbf{C} of degrees $\delta_{\alpha-1}$ and δ_α respectively, it follows that $\frac{|F^{(\alpha-1)}|^2 |F^{(\alpha)}|^2}{(1 + z\bar{z})^{\delta_{\alpha-1} + \delta_\alpha}}$ is globally defined on \mathbf{C} and has a non-zero constant limit c , as $z \rightarrow \infty$. So from (40) we get

$$\frac{|F^{(\alpha-1)}|^2 |F^{(\alpha)}|^2}{(1 + z\bar{z})^{\delta_{\alpha-1} + \delta_\alpha}} = c.$$

Then we have

$$|F^{(\alpha-1)}|^2 = c_{\alpha-1} (1 + z\bar{z})^{\delta_{\alpha-1}}, \quad |F^{(\alpha)}|^2 = c_\alpha (1 + z\bar{z})^{\delta_\alpha},$$

where $c_{\alpha-1}$ and c_α are constants.

Hence, $l_{\alpha-1} = \frac{\delta_{\alpha-1}}{(1 + z\bar{z})^2}$ and $l_\alpha = \frac{\delta_\alpha}{(1 + z\bar{z})^2}$, namely, φ_α is of constant curvature and constant Kähler angle. □

From (19) and (22) we know that if $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is a pseudo-holomorphic curve with constant Kähler angle θ_α , then $K_\alpha = \frac{1}{1 + t_\alpha} K_{(\alpha)}$.

Remark. We do not need to assume that $\varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ is non-degenerate in Theorem 4.2.

To conclude this section, we give an example. This example is a harmonic sequence, in which the Gauss curvature and the Kähler angle of each element are constant.

Let $f_0(z) = (1, 0, \sqrt{2}z, 0, z^2)$ and $g_0(z) = (0, 1, 0, z, 0)$; then

$$\varphi_0 = \frac{1}{(1 + z\bar{z})^2} \begin{pmatrix} 1 & 0 & \sqrt{2}z & 0 & z^2 \\ 0 & 1 + z\bar{z} & 0 & z(1 + z\bar{z}) & 0 \\ \sqrt{2}\bar{z} & 0 & 2z\bar{z} & 0 & \sqrt{2}z^2\bar{z} \\ 0 & \bar{z}(1 + z\bar{z}) & 0 & z\bar{z}(1 + z\bar{z}) & 0 \\ \bar{z}^2 & 0 & \sqrt{2}z\bar{z}^2 & 0 & z^2\bar{z}^2 \end{pmatrix} : \mathbf{S}^2 \rightarrow \mathbf{G}_{2,5}$$

determined by $f_0(z)$ and $g_0(z)$ is a holomorphic map.

An immediate computation shows that

$$f_1(z, \bar{z}) = \varphi_0^\perp(\partial f_0(z)) = \left(-\frac{2\bar{z}}{1+z\bar{z}}, 0, \frac{\sqrt{2}(1-z\bar{z})}{1+z\bar{z}}, 0, \frac{2z}{1+z\bar{z}} \right),$$

$$g_1(z, \bar{z}) = \varphi_0^\perp(\partial g_0(z)) = \left(0, -\frac{\bar{z}}{1+z\bar{z}}, 0, \frac{1}{1+z\bar{z}}, 0 \right),$$

and $\varphi_1(z, \bar{z})$ determined by f_1 and g_1 is given by

$$\varphi_1 = \frac{1}{(1+z\bar{z})^2} \begin{pmatrix} 2z\bar{z} & 0 & \sqrt{2}z(z\bar{z}-1) & 0 & -2z^2 \\ 0 & z\bar{z}(1+z\bar{z}) & 0 & -z(1+z\bar{z}) & 0 \\ \sqrt{2}z(z\bar{z}-1) & 0 & (z\bar{z}-1)^2 & 0 & \sqrt{2}z(z\bar{z}-1) \\ 0 & -\bar{z}(1+z\bar{z}) & 0 & 1+z\bar{z} & 0 \\ -2\bar{z}^2 & 0 & \sqrt{2}\bar{z}(z\bar{z}-1) & 0 & 2z\bar{z} \end{pmatrix},$$

which is obviously a pseudo-holomorphic curve into $\mathbf{G}_{2,5}$. Similarly, we have

$$f_2 = \varphi_1^\perp(\partial f_1) = \left(\frac{2\bar{z}^2}{(1+z\bar{z})^2}, 0, -\frac{2\sqrt{2}\bar{z}}{(1+z\bar{z})^2}, 0, \frac{2}{(1+z\bar{z})^2} \right),$$

$$g_2 = \varphi_1^\perp(\partial g_1) = (0, 0, 0, 0, 0),$$

and φ_2 , determined by f_2 and g_2 , is given by

$$\varphi_2 = \frac{1}{(1+z\bar{z})^2} \begin{pmatrix} z^2\bar{z}^2 & 0 & -\sqrt{2}z^2\bar{z} & 0 & z^2 \\ 0 & 0 & 0 & 0 & 0 \\ -\sqrt{2}z\bar{z}^2 & 0 & 2z\bar{z} & 0 & -\sqrt{2}z \\ 0 & 0 & 0 & 0 & 0 \\ \bar{z}^2 & 0 & -\sqrt{2}\bar{z} & 0 & 1 \end{pmatrix}.$$

φ_2 is an anti-holomorphic curve, which is isomorphic to the Veronese curve, in \mathbf{CP}^2 .

Hence we obtain a harmonic sequence from φ_0 :

$$0 \xrightarrow{\partial'} \varphi = \varphi_0 \xrightarrow{\partial'} \varphi_1 \xrightarrow{\partial'} \varphi_2 \xrightarrow{\partial'} 0.$$

By a straightforward computation we obtain

$$l_0 = \frac{3}{(1+z\bar{z})^2}, \quad l_1 = \frac{2}{(1+z\bar{z})^2}, \quad l_2 = 0.$$

It is very easy to see that $K(\varphi_0) = 4/3$, $K(\varphi_1) = 4/5$, $K(\varphi_2) = 2$ and $t_1 = 3/2$.

It is well known that the rigidity of holomorphic curves in Grassmannians fails; so this example is a special harmonic sequence.

5. PINCHING THEOREM FOR CURVATURE AND KÄHLER ANGLE

In this section we will discuss curvature pinching and Kähler angle pinching of non-degenerate pseudo-holomorphic spheres in complex Grassmann manifolds.

Let $\varphi = \varphi_\alpha : \mathbf{S}^2 \rightarrow \mathbf{G}_{k_\alpha, n}$ be a pseudo-holomorphic curve with the non-degenerate associated harmonic sequence (5), and let α_0 be the length of its associated harmonic sequence. Then from (28) we have

$$(41) \quad \delta_\alpha = -\delta_{\alpha-2} + 2\delta_{\alpha-1} - \tau_{\alpha-1} - 2$$

for $\alpha = 1, \dots, \alpha_0$, and

$$(42) \quad \tau_\alpha = (\delta_\alpha - \delta_{\alpha+1}) - (\delta_{\alpha-1} - \delta_\alpha) - 2$$

for $\alpha = 0, 1, \dots, \alpha_0 - 1$.

It is an immediate consequence of (41) and (42) that

$$(43) \quad \delta_\alpha = (\alpha + 1)(\delta_0 - \alpha) - \sum_{\beta=0}^{\alpha-1} (\alpha - \beta)\tau_\beta$$

for $\alpha = 1, \dots, \alpha_0$, and

$$(44) \quad \tau_0 + \dots + \tau_\alpha = (\delta_\alpha - \delta_{\alpha+1}) + \delta_0 - 2(\alpha + 1)$$

for $\alpha = 0, 1, \dots, \alpha_0 - 1$, where δ_0 is the degree of the holomorphic map φ_0 in (5).

From (43) and (44) we have also

$$(45) \quad \sum_{\alpha=0}^{\alpha_0-1} (\alpha_0 - \alpha)\tau_\alpha = (\alpha_0 + 1)(\delta_0 - \alpha_0)$$

and

$$(46) \quad \delta_\alpha = (\alpha + 1)(\alpha_0 - \alpha) + \frac{\alpha_0 - \alpha}{\alpha_0 + 1} \sum_{\beta=0}^{\alpha-1} (\beta + 1)\tau_\beta + \frac{\alpha + 1}{\alpha_0 + 1} \sum_{\beta=\alpha}^{\alpha_0-1} (\alpha_0 - \beta)\tau_\beta.$$

Denoting $\tau = \min \{\tau_0, \dots, \tau_{\alpha_0-1}\} (\geq 0)$, we immediately obtain

$$(47) \quad \delta_0 \geq \alpha_0(1 + \tfrac{1}{2}\tau), \quad \delta_\alpha \geq (\alpha + 1)(\alpha_0 - \alpha)(1 + \tfrac{1}{2}\tau),$$

and “=” holds if and only if $\tau_0 = \dots = \tau_{\alpha_0-1}$, where $\alpha = 0, 1, \dots, \alpha_0 - 1$.

Obviously, φ is a totally unramified non-degenerate pseudo-holomorphic minimal immersion, i.e., the harmonic sequence $\varphi_0, \dots, \varphi_{\alpha_0} : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ is non-degenerate and totally unramified if and only if the degree δ_0 of φ_0 is α_0 . For a totally unramified non-degenerate harmonic sequence $\varphi_0, \dots, \varphi_{\alpha_0} : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ we have

$$(48) \quad \delta_\alpha = (\alpha + 1)(\alpha_0 - \alpha).$$

At first, by using the Gauss-Bonnet theorem we have

Lemma 5.1. *Suppose that the curvature K_α of φ_α satisfies either $K_\alpha \geq \frac{4}{\delta_{\alpha-1} + \delta_\alpha}$ or $K_\alpha \leq \frac{4}{\delta_{\alpha-1} + \delta_\alpha}$. Then $K_\alpha = \frac{4}{\delta_{\alpha-1} + \delta_\alpha}$.*

Remark. In Lemma 5.1 we do not need to assume that φ_α is non-degenerate.

Theorem 5.2. *Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a pseudo-holomorphic curve with non-degenerate associated harmonic sequence, and suppose that φ is the α -th element φ_α of its non-degenerate associated harmonic sequence.*

(i) *If $K(\varphi) \geq \frac{4}{(\alpha_0 + 2\alpha(\alpha_0 - \alpha))(1 + \frac{1}{2}\tau)}$, then*

$$K(\varphi) = \frac{4}{(\alpha_0 + 2\alpha(\alpha_0 - \alpha))(1 + \frac{1}{2}\tau)},$$

and $\tau_\beta = \tau$ for all β .

(ii) *If $K(\varphi) \leq \frac{4}{(\alpha_0 + 2\alpha(\alpha_0 - \alpha))(1 + \frac{1}{2}\tau)}$ and if $\tau_\beta = \tau$ for all β , then*

$$K(\varphi) = \frac{4}{(\alpha_0 + 2\alpha(\alpha_0 - \alpha))(1 + \frac{1}{2}\tau)}.$$

Proof. From (47) we see that

$$\delta_{\alpha-1} + \delta_{\alpha} \geq (\alpha_0 + 2\alpha(\alpha_0 - \alpha))(1 + \tfrac{1}{2}\tau),$$

with equality if and only if $\tau_{\beta} = \tau$ for all β . The result is now immediate from Lemma 5.1. \square

Remark. We have $t_{\alpha} = \frac{\alpha(\alpha_0 - \alpha + 1)}{(\alpha + 1)(\alpha_0 - \alpha)}$ under the assumption of Theorem 5.2. This shows that the Kähler angle θ_{α} is independent of τ .

The following is an immediate consequence of Theorem 5.2.

Corollary 5.3. *Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a holomorphic curve with non-degenerate associated harmonic sequence. Suppose $K(\varphi) \leq \frac{4}{\alpha_0(1 + \frac{1}{2}\tau)}$ and $\tau_{\beta} = \tau$ for all β .*

Then $K(\varphi) = \frac{4}{\alpha_0(1 + \frac{1}{2}\tau)}$.

Similarly, the following theorem is also an immediate consequence of Theorem 5.2.

Corollary 5.4. *Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a holomorphic curve with non-degenerate associated harmonic sequence, and suppose $K(\varphi) \geq \frac{4}{\alpha_0(1 + \frac{1}{2}\tau)}$. Then $K(\varphi) =$*

$\frac{4}{\alpha_0(1 + \frac{1}{2}\tau)}$, and $\tau_{\beta} = \tau$ for all β .

We now prove a pinching theorem for the Kähler angle. Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a pseudo-holomorphic sphere and let $\varphi_0, \dots, \varphi_{\alpha_0}$ be the associated harmonic sequence. We assume that $\varphi = \varphi_{\alpha}$.

Lemma 5.5 ([1]). *If the Kähler angle t_{α} of φ_{α} satisfies either $t_{\alpha} \geq \frac{\delta_{\alpha-1}}{\delta_{\alpha}}$ or*

$t_{\alpha} \leq \frac{\delta_{\alpha-1}}{\delta_{\alpha}}$, then $t_{\alpha} = \frac{\delta_{\alpha-1}}{\delta_{\alpha}}$.

Lemma 5.6. *Let φ_{α} be a pseudo-holomorphic curve with non-degenerate associated harmonic sequence. If $\tau_{\beta} = \tau$ for all β , and t_{α} satisfies either $t_{\alpha} \geq \frac{\alpha(\alpha_0 - \alpha + 1)}{(\alpha + 1)(\alpha_0 - \alpha)}$*

or $t_{\alpha} \leq \frac{\alpha(\alpha_0 - \alpha + 1)}{(\alpha + 1)(\alpha_0 - \alpha)}$, then $t_{\alpha} = \frac{\alpha(\alpha_0 - \alpha + 1)}{(\alpha + 1)(\alpha_0 - \alpha)}$.

The proof of the above theorem is immediate from Lemma 5.5 and (47).

Using (46), we can also prove the following.

Theorem 5.7. *Let $\varphi : \mathbf{S}^2 \rightarrow \mathbf{G}_{k,n}$ be a pseudo-holomorphic curve with non-degenerate associated harmonic sequence, and suppose that φ is the α -th element φ_{α} of its non-degenerate associated harmonic sequence. If $t_{\alpha} \leq \frac{1}{2}$ (resp. $t_{\alpha} \geq 2$), then $t_{\alpha} = 0$ (resp. $t_{\alpha} = \infty$), i.e., φ is a holomorphic (resp. anti-holomorphic) curve.*

Proof. When $\alpha \neq 0$ and α_0 , by (46) an immediate computation shows that

$$\frac{1}{2} < \frac{\delta_{\alpha-1}}{\delta_{\alpha}} < 2.$$

Hence, by Lemma 5.5, if $t_{\alpha} \leq \frac{1}{2}$ (resp. $t_{\alpha} \geq 2$), then $\alpha = 0$ (resp. $\alpha = \alpha_0$), i.e., φ is a holomorphic (resp. anti-holomorphic) curve. \square

We believe that $\tau \neq 0$ for the non-degenerate harmonic sequence associated to the holomorphic curve of constant curvature, except for the Veronese sequence.

REFERENCES

- [1] J. Bolton, G. R. Jensen, M. Rigoli, and L. M. Woodward, *On conformal minimal immersions of S^2 into CP^n* , Math. Ann., 279(1988), 599-620. MR **88m**:53110
- [2] E. Calabi, *Isometric embedding of complex manifolds*, Ann. Math. (2), 58(1953), 1-23. MR **15**:160c
- [3] S. S. Chern and J. G. Wolfson, *Harmonic maps of the two-sphere into a complex Grassmann manifold, II*, Ann. Math., 125(1987), 301-335. MR **88g**:58038
- [4] S. S. Chern and J. G. Wolfson, *Minimal surfaces by moving frames*, Amer. J. Math., 105(1983), 59-83. MR **84i**:53056
- [5] Q. Chi and Y. Zheng, *Rigidity of pseudo-holomorphic curves of constant curvature in Grassmann manifolds*, Trans. Amer. Math. Soc., 313(1989), 393-406. MR **90m**:53072
- [6] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Pure and Applied Mathematics, London, New York: Wiley, 1978. MR **80b**:14001
- [7] X. X. Jiao, *On harmonic maps of surfaces into complex Grassmannians*, Chinese Ann. Math., 21A(1)(2000), 57-60. MR **2001b**:53083
- [8] H. B. Lawson, *The Riemannian geometry of holomorphic curves*, Proc. Conf. Holomorphic Mapping and Minimal Surfaces, Bol. Soc. Brasil. Mat., 2(1971), 45-62. MR **48**:2957
- [9] K. Uhlenbeck, *Harmonic maps into Lie groups (classical solutions of the chiral model)*, J. Differential Geom., 30(1989), 1-50. MR **90g**:58028
- [10] J. G. Wolfson, *Harmonic sequences and harmonic maps of surfaces into complex Grassmann manifolds*, J. Differential Geom., 27(1988), 161-178. MR **89c**:58031
- [11] Y. B. Zheng, *Quantization of curvature of harmonic two-spheres in Grassmann manifolds*, Trans. Amer. Math. Soc., 316(1)(1989), 193-214. MR **90b**:58055
- [12] K. Yang, *Complete and compact minimal surfaces*, Kluwer Academic Publishers, 1989. MR **91h**:53058
- [13] K. Yang, *Compact Riemann surfaces and algebraic curves*, Series in Pure Mathematics, Vol. 10, World Scientific, 1988. MR **90e**:14023
- [14] X. X. Jiao and J. G. Peng, *A classification of holomorphic two-spheres with constant curvature in complex Grassmannians*, Differential Geom. Appl., to appear.

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL, CHINESE ACADEMY OF SCIENCES, BEIJING 100039, CHINA

E-mail address: `xxj@gscas.ac.cn`

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL, CHINESE ACADEMY OF SCIENCES, BEIJING 100039, CHINA

E-mail address: `pengck@gscas.ac.cn`

THE PERIODIC EULER-BERNOULLI EQUATION

VASSILIS G. PAPANICOLAOU

ABSTRACT. We continue the study of the Floquet (spectral) theory of the beam equation, namely the fourth-order eigenvalue problem

$$[a(x)u''(x)]'' = \lambda\rho(x)u(x), \quad -\infty < x < \infty,$$

where the functions a and ρ are periodic and strictly positive. This equation models the transverse vibrations of a thin straight (periodic) beam whose physical characteristics are described by a and ρ . Here we develop a theory analogous to the theory of the Hill operator $-(d/dx)^2 + q(x)$.

We first review some facts and notions from our previous works, including the concept of the pseudospectrum, or ψ -spectrum.

Our new analysis begins with a detailed study of the zeros of the function $F(\lambda; k)$, for any given “quasimomentum” $k \in \mathbb{C}$, where $F(\lambda; k) = 0$ is the Floquet-Bloch variety of the beam equation (the Hill quantity corresponding to $F(\lambda; k)$ is $\Delta(\lambda) - 2\cos(kb)$, where $\Delta(\lambda)$ is the discriminant and b the period of q). We show that the multiplicity $m(\lambda^*)$ of any zero λ^* of $F(\lambda; k)$ can be one or two and $m(\lambda^*) = 2$ (for some k) if and only if λ^* is also a zero of another entire function $D(\lambda)$, independent of k . Furthermore, we show that $D(\lambda)$ has exactly one zero in each gap of the spectrum and two zeros (counting multiplicities) in each ψ -gap. If λ^* is a double zero of $F(\lambda; k)$, it may happen that there is only one Floquet solution with quasimomentum k ; thus, there are exceptional cases where the algebraic and geometric multiplicities do not agree.

Next we show that if (α, β) is an open ψ -gap of the pseudospectrum (i.e., $\alpha < \beta$), then the Floquet matrix $T(\lambda)$ has a specific Jordan anomaly at $\lambda = \alpha$ and $\lambda = \beta$.

We then introduce a multipoint (Dirichlet-type) eigenvalue problem which is the analogue of the Dirichlet problem for the Hill equation. We denote by $\{\mu_n\}_{n \in \mathbb{Z}}$ the eigenvalues of this multipoint problem and show that $\{\mu_n\}_{n \in \mathbb{Z}}$ is also characterized as the set of values of λ for which there is a proper Floquet solution $f(x; \lambda)$ such that $f(0; \lambda) = 0$.

We also show (Theorem 7) that each gap of the $L^2(\mathbb{R})$ -spectrum contains exactly one μ_n and each ψ -gap of the pseudospectrum contains exactly two μ_n 's, counting multiplicities. Here when we say “gap” or “ ψ -gap” we also include the endpoints (so that when two consecutive bands or ψ -bands touch, the in-between collapsed gap, or ψ -gap, is a point). We believe that $\{\mu_n\}_{n \in \mathbb{Z}}$ can be used to formulate the associated inverse spectral problem.

As an application of Theorem 7, we show that if ν^* is a collapsed (“closed”) ψ -gap, then the Floquet matrix $T(\nu^*)$ is diagonalizable.

Some of the above results were conjectured in our previous works. However, our conjecture that if all the ψ -gaps are closed, then the beam operator is the square of a second-order (Hill-type) operator, is still open.

Received by the editors November 13, 2001 and, in revised form, November 10, 2002.

2000 *Mathematics Subject Classification.* Primary 34B05, 34B10, 34B30, 34L40, 74B05.

Key words and phrases. Euler-Bernoulli equation for the vibrating beam, beam operator, Hill operator, Floquet spectrum, pseudospectrum, algebraic/geometric multiplicity, multipoint eigenvalue problem.

1. INTRODUCTION

The term “periodic Euler-Bernoulli equation” refers to the eigenvalue problem

$$(1) \quad [a(x)u''(x)]'' = \lambda \rho(x)u(x), \quad -\infty < x < \infty,$$

where $a(x)$ and $\rho(x)$ are strictly positive and periodic with a common period b , satisfying the smoothness conditions $a \in C^2(\mathbb{R})$ and $\rho \in C(\mathbb{R})$. Furthermore, without loss of generality, $a(x)$ and $\rho(x)$ are normalized so that

$$(2) \quad \int_0^b \left[\frac{\rho(x)}{a(x)} \right]^{1/4} dx = b.$$

One advantage of this normalization is that the asymptotics of certain quantities, such as $|\lambda| \rightarrow \infty$, become simpler, and this is the only reason that (2) is used in the present work (see Section 3).

The study of (1) was initiated by the author in [31] and [33]. There are theoretical as well as practical reasons for studying (1). The Floquet (spectral) theory of (1) is richer than its second-order counterpart (namely the Sturm-Liouville problem with periodic coefficients, also known as Hill's equation, or the one-dimensional Schrödinger equation with a periodic potential). All the main second-order properties continue to hold, while new interesting phenomena arise which are nonexistent in the second-order case. On the practical side, we notice that a typical application of (1) is that it models the transverse vibrations of a thin straight beam with periodic characteristics (see, e.g., [36] or [17]). Elastic structures consisting of many thin elements arranged periodically are quite common. Although there are some authors who have studied such problems numerically (see, for example, [28]), as far as we know, [31] and [33] are the only theoretical works on (1). However, recently there is an increasing interest in higher-order periodic eigenvalue problems (e.g., [2], [6], [7]). For results on the second-order inverse periodic problem, or higher-order nonperiodic inverse problems, the reader may see, e.g., [3], [4], [5], [8], [10], [14], [15], [17], [23], [24], [25], [26], [27], [32], [35], [38].

The present work continues the investigation on the Floquet spectral theory of (1), initiated in [31] and [33]. The goal is a theory analogous to the theory of the Hill operator $-(d/dx)^2 + q(x)$.

In Section 2 we review some facts and notions from our previous works, including the concept of the pseudospectrum, or ψ -spectrum, introduced in [33]. Theorems numbered here by capital Latin letters have been proved in our previous works [31] and [33]. At the end of Section 2 we have included a subsection containing some ideas on the significance of the pseudospectrum.

The new analysis begins in Section 3. We first describe (in Subsection 3.1) a technique we use for proving certain theorems (Theorems 2 and 7 of this work). Then, in Subsection 3.2 we give a detailed study of the zeros of the function $F(\lambda; k)$, in the spirit of [16], for any given “quasimomentum” $k \in \mathbb{C}$, where $F(\lambda; k) = 0$ is the Floquet-Bloch variety of the beam equation (the Hill quantity corresponding to $F(\lambda; k)$ is $\Delta(\lambda) - 2 \cos(kb)$, where $\Delta(\lambda)$ is the discriminant of the Hill operator and b the period of q). We show that the multiplicity $m(\lambda^*)$ of any zero λ^* of $F(\lambda; k)$ can be one or two and $m(\lambda^*) = 2$ (for some k) if and only if λ^* is also a zero of another entire function $D(\lambda)$, independent of k . Furthermore, we show that $D(\lambda)$ has exactly one zero in each gap of the spectrum and two zeros (counting multiplicities) in each ψ -gap. If λ^* is a double zero of $F(\lambda; k)$ it may happen

that there is only one Floquet solution with quasimomentum k ; thus, there are exceptional cases where the algebraic and geometric multiplicities do not agree.

In Section 4 we first (Subsection 4.1) review briefly some facts regarding certain operators L_k , where $k \in \mathbb{C}$ is the quasimomentum, and then, in Subsection 4.2, we apply them to show that if (α, β) is an open ψ -gap of the pseudospectrum (i.e., $\alpha < \beta$), then the Floquet matrix $T(\lambda)$ has a specific Jordan anomaly at $\lambda = \alpha$ and $\lambda = \beta$ (this was conjectured in [31] and [33]).

In Section 5 we introduce a multipoint (Dirichlet-type) eigenvalue problem which is the analogue of the Dirichlet problem for the Hill equation. We denote by $\{\mu_n\}_{n \in \mathbb{Z}}$ the eigenvalues of this multipoint problem and show that $\{\mu_n\}_{n \in \mathbb{Z}}$ is also characterized as the set of values of λ for which there is a proper Floquet solution $f(x; \lambda)$ such that $f(0; \lambda) = 0$. If we normalize f so that $f(0; \lambda) = 1$, then $\{\mu_n\}_{n \in \mathbb{Z}}$ is the set of poles of $f(x; \lambda)$ (viewed as a function of λ , of course) counting multiplicities (this approach is used in [11]).

We also show (Theorem 7) that each gap of the $L^2(\mathbb{R})$ -spectrum contains exactly one μ_n and each ψ -gap of the pseudospectrum contains exactly two μ_n 's, counting multiplicities. Here when we say "gap" or " ψ -gap" we also include the endpoints (so that when two consecutive bands or ψ -bands touch, the in-between collapsed gap, or ψ -gap, is a point). We believe that $\{\mu_n\}_{n \in \mathbb{Z}}$ can be used to formulate the associated inverse spectral problem.

As an application of Theorem 7, we show that if ν^* is a collapsed ("closed") ψ -gap, then the Floquet matrix $T(\nu^*)$ is diagonalizable (this too was conjectured in [31] and [33]).

Some of the above results were conjectured in our previous works. However, the formulation of the inverse problem (and, in particular, our conjecture that if all the ψ -gaps are closed, then the beam operator is the square of a second-order, Hill-type, operator) remains open.

2. REVIEW OF EARLIER RESULTS

2.1. The spectrum. We start by recalling certain general facts related to (1) (see [13], Sec. XIII.7) and some of the main results established in [31] and [33] (other references for Floquet or periodic spectral theory are, e.g., [9], [10], [16], [18], [19], [22], [34]). The problem is selfadjoint (with no boundary conditions at $\pm\infty$). The underlying operator L (the "Euler-Bernoulli operator" or "beam operator") is given by

$$Lu = \rho^{-1} (au'')''.$$

The corresponding Hilbert space is the ρ -weighted space $L^2_\rho(\mathbb{R})$. Notice that L is a product of two second-order differential operators, namely $L = L_2 L_1$, where $L_1 u = -au''$ and $L_2 u = -\rho^{-1} u''$.

For any fixed λ the "shift" transformation

$$(Tu)(x) = u(x + b)$$

maps solutions of (1) to solutions. As a basis of this space we take the solutions $u_j(x; \lambda)$, $j = 1, 2, 3, 4$, such that (primes refer to derivatives with respect to x ; δ_{jk} is the Kronecker delta)

$$u_j^{(k-1)}(0; \lambda) = \delta_{jk}, \quad k = 1, 2, \quad a(0)u_j''(0; \lambda) = \delta_{j3}, \quad [au_j']'(0; \lambda) = \delta_{j4}.$$

We refer to u_j as the j -th fundamental solution. Each $u_j(x; \lambda)$ is entire in λ of order $1/4$. We identify T with its matrix representation with respect to the above basis (called Floquet matrix), namely

$$T = \begin{bmatrix} u_1(b) & u_2(b) & u_3(b) & u_4(b) \\ u'_1(b) & u'_2(b) & u'_3(b) & u'_4(b) \\ a(b)u''_1(b) & a(b)u''_2(b) & a(b)u''_3(b) & a(b)u''_4(b) \\ [au''_1]'(b) & [au''_2]'(b) & [au''_3]'(b) & [au''_4]'(b) \end{bmatrix},$$

where the dependence in λ is suppressed for typographical convenience. In [31] it was shown that the eigenvalues r_1, r_2, r_3, r_4 of T (called Floquet multipliers) appear in pairs of inverses, namely

$$(3) \quad r_1 r_4 = r_2 r_3 = 1$$

(in fact this is true for any selfadjoint ordinary differential operator with real, periodic coefficients). It follows that the characteristic equation of T has the form

$$(4) \quad r^4 - A(\lambda)r^3 + \tilde{B}(\lambda)r^2 - A(\lambda)r + 1 = 0.$$

Except for a discrete set of λ 's, $T = T(\lambda)$ is similar to a diagonal matrix and its eigenvectors correspond to the Floquet solutions, namely to the solutions f_j , $j = 1, 2, 3, 4$ of (1) such that

$$(5) \quad f_j(x + b) = r_j f_j(x).$$

There are four linearly independent Floquet solutions if and only if T is similar to a diagonal matrix. We also notice that (5) implies

$$f_j(x) = e^{w_j x} p_j(x), \quad \text{where } r_j = e^{w_j b} \quad \text{and} \quad p_j(x + b) = p_j(x).$$

The $L^2_\rho(\mathbb{R})$ -spectrum $S(a, \rho)$ of (1) can be characterized as the set

$$(6) \quad S(a, \rho) = \{\lambda \in \mathbb{C} : |r_j(\lambda)| = 1, \text{ for some } j\}.$$

It can be shown that $S(a, \rho)$ is real with $\inf S(a, \rho) = 0$. In the unperturbed case, i.e., when $a(x) \equiv \rho(x) \equiv 1$, we have

$$S_0 \stackrel{\text{def}}{=} S(1, 1) = [0, \infty)$$

(the index 0 indicates that a quantity belongs to the unperturbed case).

Next, for a fixed real number k , we consider the corresponding k -Floquet eigenvalue problem on $(0, b)$, namely

$$(7) \quad [a(x)u''(x)]'' = \lambda \rho(x)u(x), \quad u^{(j)}(b) = e^{ikb}u^{(j)}(0), \quad j = 0, 1, 2, 3.$$

Let $\{\lambda_n(k)\}_{n=1}^\infty$ be the spectrum of (7). Since the problem is selfadjoint, $\lambda_n(k) \in \mathbb{R}$. The eigenvalues can be indexed so that $\lambda_n(k) \leq \lambda_{n+1}(k)$. Then $\lambda_n(k) \sim Cn^4$. We also notice that, since $\lambda_n(k + 2\pi/b) = \lambda_n(k)$, one only needs to consider k in $[0, 2\pi/b)$. Furthermore, (3) implies that $\lambda_n(2\pi/b - k) = \lambda_n(k)$. The set $\{\lambda_{n-1} \stackrel{\text{def}}{=} \lambda_n(0)\}_{n=1}^\infty$ is the periodic spectrum, while $\{\lambda'_n \stackrel{\text{def}}{=} \lambda_n(\pi/b)\}_{n=1}^\infty$ is the antiperiodic spectrum. The n -th band of $S(a, \rho)$ is the closed interval

$$B_n = \bigcup_{0 \leq k \leq \pi/b} \lambda_n(k)$$

and it is well known that $S(a, \rho) = \bigcup_{n=1}^{\infty} B_n$. In [31] it was shown that (as in the Hill equation)

$$B_{2m+1} = [\lambda_{2m}, \lambda'_{2m+1}], \quad B_{2m+2} = [\lambda'_{2m+2}, \lambda_{2m+1}], \quad m = 0, 1, 2, \dots$$

In fact, as λ moves, say with constant velocity, from λ_{2m} to λ'_{2m+1} (resp. from λ'_{2m+2} to λ_{2m+1}) two Floquet multipliers, say r_2 and $r_3 = \bar{r}_2$, move smoothly on the unit circle, with nonvanishing speed (i.e., without changing direction), starting at 1 and arriving at -1 (resp. starting at -1 and arriving at 1). The other two multipliers, r_1 and r_4 , stay real. In particular, $\lambda_n(k)$ is strictly monotone in k , on $[0, \pi/b]$; hence the interior of B_n is never empty. An interesting question here (pointed out by the anonymous referee) is whether we always have $|d\lambda_n/dk| > 1$, as in the Hill case (see [19]).

Two bands can “touch” each other (when $\lambda_{2m+1} = \lambda_{2m+2}$ or $\lambda'_{2m+1} = \lambda'_{2m+2}$), but they cannot overlap (i.e., they have disjoint interiors). The gaps of the spectrum $S(a, \rho)$ are

$$I_{2m-1} = (\lambda'_{2m-1}, \lambda'_{2m}), \quad I_{2m} = (\lambda_{2m-1}, \lambda_{2m}), \quad m = 1, 2, 3, \dots,$$

and empty gaps are traditionally called “closed”. If $\lambda_{2m-1} < \lambda_{2m}$ (resp. $\lambda'_{2m-1} < \lambda'_{2m}$), then r_2 (and r_3) has square root branch points at $\lambda = \lambda_{2m-1}, \lambda_{2m}$ (resp. at $\lambda = \lambda'_{2m-1}, \lambda'_{2m}$). If, on the other hand, $\lambda_{2m-1} = \lambda_{2m}$ (resp. $\lambda'_{2m-1} = \lambda'_{2m}$), i.e., if the corresponding gap is closed, then r_2 (and r_3) are analytic about $\lambda = \lambda_{2m-1}$ (resp. about $\lambda = \lambda'_{2m-1}$). The value $\lambda = \lambda_0 = 0$ is very special since all Floquet multipliers have a fourth root branch point there and T has only one eigenvector.

If $\lambda \neq 0$, then the characteristic equation of T can only have simple or double roots. Now let $\lambda \neq 0$ be such that (4) has a double root, say r_j . Then there is one Floquet solution $f_j(x+b) = r_j f_j(x)$ and a solution $g_j(x)$ (f_j and g_j are linearly independent) such that $g_j(x+b) = r_j g_j(x) + c_j f_j(x)$, where the constant c_j may be 0 (in this case we say that we have coexistence, i.e., two linearly independent Floquet solutions corresponding to the same multiplier). If $c_j \neq 0$, we can say that, for this particular λ , T has a Jordan anomaly (this terminology is due to Professor Barry Simon) and that $g_j(x)$ is a generalized Floquet solution of (1).

We now review the main results of [33]. Notice that, in that reference it was assumed that $a, \rho \in C^4(\mathbb{R})$, but we believe that they remain true for $a \in C^2(\mathbb{R})$ and $\rho \in C(\mathbb{R})$, and here is why: All these results concern entire functions of λ that are polynomial expressions of the $u_j(b; \lambda)$'s. By considering $a(x)$ and $\rho(x)$ as limits of smooth (C^4 or even C^∞) functions $a_n(x), \rho_n(x)$ in the C^2 - and sup-norms respectively, the corresponding entire function $u_{j,n}(b; \lambda)$ converges to $u_j(b; \lambda)$, uniformly on compact subsets of \mathbb{C} (see the proposition in the Appendix), for $j = 1, 2, 3, 4$. Thus, we think that our results extend immediately to less smooth coefficients.

For a fixed real number k , equation (4) implies (by setting $r = e^{ikb}$) that the k -Floquet eigenvalues of (7) are the zeros of the entire function

$$(8) \quad F(\lambda; k) = B(\lambda) - 2A(\lambda) \cos(kb) + 4 \cos^2(kb),$$

where we have set

$$(9) \quad B(\lambda) = \tilde{B}(\lambda) - 2.$$

In the unperturbed case, this function becomes

$$(10) \quad F_0(\lambda; k) = 4 \left[\cosh(\lambda^{1/4}b) - \cos(kb) \right] \left[\cos(\lambda^{1/4}b) - \cos(kb) \right].$$

This expression implies easily that, if $0 < k < \pi/b$, then the zeros of $F_0(\lambda; k)$ are

$$(11) \qquad \lambda_{n,0} = \left[2 \lfloor n/2 \rfloor \pi/b - (-1)^n k \right]^4, \qquad n = 1, 2, 3, \dots$$

(where $\lfloor \cdot \rfloor$ is the greatest integer function) and they are all simple. Furthermore, $\lambda_{n,0}$ lies in the interior of the n -th band $B_{n,0}$, for every n .

(Theorems that have been proved in previous articles are numbered by upper-case Roman letters.)

Theorem A. *Let $0 < k < \pi/b$. Then the zeros $\{\lambda_n(k)\}_{n=1}^\infty$ of $F(\lambda; k)$ are simple, $\lambda_n(k) \in B_n$, and to each $\lambda_n(k)$ corresponds a unique eigenfunction $\phi_n(x; k)$ of (7), i.e., the geometric multiplicity of $\lambda_n(k)$ is also 1.*

Theorem B. *The multiplicity of any zero of $F^+(\lambda) \stackrel{\text{def}}{=} F(\lambda; 0)$ can be only 1 or 2. A zero λ^* of $F^+(\lambda)$ is double if and only if $\lambda^* = \lambda_{2m-1} = \lambda_{2m}$, for some $m \geq 1$ (i.e., the corresponding gap is closed). Furthermore, (7) has two (linearly independent) periodic eigenfunctions corresponding to λ^* (coexistence) if and only if $\lambda^* = \lambda_{2m-1} = \lambda_{2m}$. Thus, we can say that the algebraic multiplicity of any periodic eigenvalue is equal to its geometric multiplicity. The same things also hold for $F^-(\lambda) \stackrel{\text{def}}{=} F(\lambda; \pi/b)$, which is the entire function associated to the antiperiodic case.*

2.2. The pseudospectrum. The previous results are the exact analogues of the results for the Hill equation regarding algebraic and geometric multiplicities (see [16]).

In [33] we introduced a concept that, as far as we know, does not have a counterpart in the second-order case:

Definition. Let $k \in (0, \pi/b)$. The set

$$\Psi_k(a, \rho) = \{ \lambda \in \mathbb{C} : \text{there are two Floquet multipliers } r_j(\lambda), r_l(\lambda) \text{ of (1)} \\ \text{such that } r_j = \overline{r_l} = |r_j| e^{ikb}, \quad |r_j| > 1 \}$$

is called the k -Floquet pseudospectrum (or k -Floquet ψ -spectrum) of (1). We, furthermore, call the set

$$\Psi(a, \rho) = \overline{\bigcup_{0 < k < \pi/b} \Psi_k(a, \rho)}$$

the pseudospectrum (ψ -spectrum) of (1) on the line (here \overline{D} denotes the topological closure of D).

The following entire function was introduced in [33]:

$$(12) \qquad G(\lambda; k) = A(\lambda)^2 - 4B(\lambda) \cos^2(kb) - 16 \cos^2(kb) \sin^2(kb).$$

It follows that, if $\nu \in \Psi_k(a, \rho)$, then $G(\nu; k) = 0$.

Since

$$G(\lambda; k) = G(\lambda; \pi/b - k),$$

the zeros of $G(\lambda; k)$ also include $\Psi_{\pi/b-k}(a, \rho)$.

Conversely, for a fixed $k \in (0, \pi/2b)$, let ν be a zero of $G(\lambda; k)$. Then (12) implies that

$$(13) \qquad A(\nu)^2 - 4B(\nu) \cos^2(kb) - 16 \cos^2(kb) \sin^2(kb) = 0.$$

Now by (3), (4), and (9),

$$(14) \quad A = r_1 + \frac{1}{r_1} + r_2 + \frac{1}{r_2} \quad \text{and} \quad B = \left(r_1 + \frac{1}{r_1}\right) \left(r_2 + \frac{1}{r_2}\right);$$

hence, given r_1 , (13) becomes a 4-th degree (algebraic) equation in r_2 . One can check that its solutions are

$$r_2 = r_1 e^{\pm 2ikb} \quad \text{and} \quad r_2 = r_1^{-1} e^{\pm 2ikb}.$$

This means that

$$\nu \in \Psi_k(a, \rho) \cup \Psi_{\pi/b-k}(a, \rho) = \{\lambda : r_j = r_l e^{\pm 2ikb}, \quad |r_j| \neq 1\}.$$

Therefore, for $k \in (0, \pi/b)$, $k \neq \pi/(2b)$, the set of zeros of $G(\lambda; k)$ is $\Psi_k(a, \rho) \cup \Psi_{\pi/b-k}(a, \rho)$.

The value $k = \pi/(2b)$ is somehow special since

$$(15) \quad G(\lambda; \pi/(2b)) = A(\lambda)^2.$$

It follows that, if ν is a zero of $G(\lambda; \pi/(2b))$, then

$$r_2(\nu) = -r_1(\nu) \quad \text{or} \quad r_2 = -r_1(\nu)^{-1}.$$

Hence

$$\nu \in \Psi_{\pi/2b}(a, \rho) = \{\lambda : r_j = -r_l, \quad |r_j| \neq 1\}$$

(notice that, in this case all Floquet multipliers are pure imaginary). These results are implicitly contained in [33], where the following theorem was established:

Theorem C. *Let $0 < k < \pi/b$. If $k \neq \pi/(2b)$ the zeros of the entire function $G(\lambda; k)$, defined in (12), are all real, strictly negative, and simple. The zeros of $G(\lambda; \pi/(2b))$ are all real strictly negative and double. Each zero $\nu_n(k)$ of $G(\lambda; k)$ is (as a function of k) strictly monotone on the interval $(0, \pi/(2b))$ and on $(\pi/(2b), \pi/b)$.*

The function $G(\lambda; k)$, as defined in (12), makes sense for all $k \in \mathbb{C}$ (in fact, it is entire in λ, k). In particular,

$$(16) \quad E(\lambda) \stackrel{\text{def}}{=} G(\lambda; 0) = A(\lambda)^2 - 4B(\lambda).$$

This function was introduced in [31]. It was shown there that 0 is always a simple zero of $E(\lambda)$ and that $\nu \neq 0$ is a zero of $E(\lambda)$ if and only if $r_j(\nu) = r_l(\nu) \neq \pm 1$, $j \neq l$. In [33] the following theorem was proved:

Theorem D. *The nonzero zeros of $E(\lambda)$ are all real, strictly negative, and simple or double. If we denote them by*

$$0 = \nu_0 > \nu'_1 \geq \nu'_2 > \nu_1 \geq \nu_2 > \nu'_3 \geq \nu'_4 > \dots,$$

we have a pseudoband-pseudogap structure on the negative λ -axis. Each ψ -band $[\nu_0, \nu'_1]$, $[\nu'_2, \nu_1]$, $[\nu_2, \nu'_3]$, ... contains exactly one point of the ψ -spectrum $\Psi_k(a, \rho)$, for any fixed $k \in (0, \pi/b)$.

Remark 1. Since 0 is always a simple zero of $E(\lambda)$ and in the unperturbed case we have

$$(17) \quad E_0(\lambda) = 4 \left[\cosh(\lambda^{1/4}b) - \cos(\lambda^{1/4}b) \right]^2,$$

it follows by Theorem D and a simple continuity argument that $E(\lambda) > 0$, if $\lambda > 0$ or if λ is in the interior of a ψ -gap; whereas $E(\lambda) < 0$, if λ is in the interior of a

ψ -band. At the zeros ν_n, ν'_n , $n \neq 0$, of $E(\lambda)$, there are Floquet multipliers r_j, r_l , $j \neq l$, such that $r_j(\nu_n) = r_l(\nu_n) > 1$ and $r_j(\nu'_n) = r_l(\nu'_n) < -1$.

If $a(x)\rho(x) \equiv 1$ (in this case the beam operator is a "perfect square"), then all the nonzero zeros of $E(\lambda)$ are double, i.e., $\lambda E(\lambda)$ is the square of an entire function. Equivalently, all ψ -gaps are closed, i.e., empty.

For the unperturbed case we have

$$G_0(\lambda; k) = 4 \left\{ \cos \left[\lambda^{1/4} b (1 + i) \right] - \cos 2kb \right\} \left\{ \cos \left[\lambda^{1/4} b (1 - i) \right] - \cos 2kb \right\}.$$

The zeros of $G_0(\lambda; k)$ are $\nu_{n,0}(k)$ and $\nu_{n,0}(\pi/b - k)$, $n = 1, 2, 3, \dots$, where

$$\nu_{n,0}(k) = -4 \left[2 \lfloor n/2 \rfloor \pi/b - (-1)^n k \right]^4.$$

Next we set

$$\nu_{n-1,0} = \nu_{n,0}(0) = \lim_{k \searrow 0} \nu_{n,0}(k), \quad \nu'_{n,0} = \nu_{n,0}(\pi/b) = \lim_{k \nearrow \pi/b} \nu_{n,0}(k).$$

Thus, $\nu_{0,0} = 0$ and, for $m = 1, 2, 3, \dots$,

$$(18) \quad \begin{aligned} \nu_{2m-1,0} &= \nu_{2m,0} = -4 \left(\frac{2m\pi}{b} \right)^4, \\ \nu'_{2m-1,0} &= \nu'_{2m,0} = -4 \left[\frac{(2m-1)\pi}{b} \right]^4. \end{aligned}$$

These numbers are the zeros of $E_0(\lambda)$.

2.2.1. The significance of the pseudospectrum. The purpose of this short (sub)section is to elucidate certain things regarding the concept of the pseudospectrum.

Let L be an n -th order (ordinary) differential operator with periodic coefficients. Then one can consider the Floquet multipliers $r_j(\lambda)$, $j = 1, \dots, n$, of L and the corresponding Floquet solutions $f_j(x; \lambda)$, $j = 1, \dots, n$. The $r_j(\lambda)$'s are, in fact, the branches of a (multivalued) analytic function which we denote by $r(\lambda)$ and, similarly, the $f_j(x; \lambda)$'s are the branches of a λ -analytic function $f(x; \lambda)$.

Let Γ be the Riemann surface of $r(\lambda)$. If we normalize $f(x; \lambda)$ so that $f(0; \lambda) = 1$, then $f(x; \lambda)$ becomes a meromorphic function on Γ , whose set of poles we denote by $\{\mu_n\}$.

In [11] it is suggested that the (periodic) inverse spectral data for L is the Riemann surface Γ together with the set of poles $\{\mu_n\}$ (notice that each μ_n is a point on Γ , i.e., μ_n is not just a complex number, since it also contains the information: on which sheet of Γ the pole lies). This is, of course, inspired by the inverse theory of the Hill operator (see, e.g., [11], [14], [24], [25], [38]).

If L is the Euler-Bernoulli operator, the multiplier $r(\lambda)$ has two types of branch points (the point $\lambda = 0$ is special and can be considered as being of both types). The branch points of the first type lie on the positive real axis and are the endpoints of the bands of the $L^2_\rho(\mathbb{R})$ -spectrum (they are also the periodic and antiperiodic eigenvalues), exactly as in the Hill case. The branch points of the second type lie on the negative real axis and they, too, define a band-gap structure which we have called *pseudospectrum* (we have called its bands and gaps " ψ -bands" and " ψ -gaps" respectively to distinguish them from the spectral bands and gaps).

Each gap of the spectrum contains exactly one μ_n . But now there are μ_n 's that do not lie in any spectral gap. It turns out that each ψ -gap of the pseudospectrum

contains exactly *two* of those μ_n 's, counting multiplicities. The exact statement is Theorem 7 of Section 5, which is, perhaps, the crux of this work.

In conclusion, for the Euler-Bernoulli case, **we need both the $L^2_\rho(\mathbb{R})$ -spectrum and the pseudospectrum in order to determine the Riemann surface Γ and the intervals in which the μ_n 's are confined.** This is why we believe that the pseudospectrum plays an essential role in the Euler-Bernoulli inverse spectral theory. In particular, we have conjectured that if the pseudospectrum has no gaps (i.e., if it is the interval $(-\infty, 0]$), then the Euler-Bernoulli operator is a perfect square of a Hill-type operator.

3. THE ZEROS OF THE FUNCTION $F(\lambda; k)$, FOR COMPLEX k

3.1. The technique. There is a technique that we have employed for proving some of our statements regarding (1), especially properties of quantities that depend on (or are related to) the spectral parameter λ (e.g., this technique has been already used for proving Theorems C and D mentioned above). It combines continuity arguments and large $|\lambda|$ asymptotics.

Here is how the technique works: We first check that the property we want to establish holds in the unperturbed case $a(x) \equiv \rho(x) \equiv 1$. Then we deform $a(x)$ and $\rho(x)$ continuously until we reach the general case, making sure that the property remains valid (a kind of "continuous induction"). For example, we can specify the (obviously continuous) deformation

$$(19) \quad a(x; t) = ta(x) + (1 - t), \quad \rho(x; t) = c(t)^4 [t\rho(x) + (1 - t)], \quad t \in [0, 1],$$

where

$$c(t) = \frac{b}{\int_0^b \left[\frac{t\rho(x) + (1-t)}{ta(x) + (1-t)} \right]^{1/4} dx}.$$

Notice that

$$a(x; 0) \equiv \rho(x; 0) \equiv 1, \quad a(x; 1) = a(x) \quad \text{and} \quad \rho(x; 1) = \rho(x).$$

Also, since $a(x)$ and $\rho(x)$ are strictly positive, it follows that there is a constant δ_0 , independent of t , such that

$$a(x; t), \rho(x; t) \geq \delta_0 > 0, \quad \text{for all } t \in [0, 1],$$

where $a(x; t)$ and $\rho(x; t)$ satisfy the normalization condition (2), namely

$$\int_0^b \left[\frac{\rho(x; t)}{a(x; t)} \right]^{1/4} dx = b, \quad \text{for all } t \in [0, 1].$$

Finally, as functions of x , $a(x; t)$ and $\rho(x; t)$ are b -periodic and as smooth as $a(x)$ and $\rho(x)$ respectively.

In some cases, we first prove the desired result for $a(x)$ and $\rho(x)$ that are sufficiently smooth, say $a, \rho \in C^4(\mathbb{R})$, and then extend it to the more general case $a \in C^2(\mathbb{R})$ and $\rho \in C(\mathbb{R})$ by approximating $a(x)$ and $\rho(x)$ by smooth functions (in the C^2 - and sup-norms respectively). The fact that a λ -quantity of the smooth case approaches (uniformly on compact subsets of \mathbb{C}) the corresponding λ -quantity of the more general case usually follows from standard Gronwall-type estimates (e.g., see the proposition in the Appendix).

One main reason that we need smooth $a(x)$ and $\rho(x)$ is that, if this is the case, there is a Liouville-type transformation (found in [3]) that transforms (1) to a canonical fourth-order eigenvalue equation, namely

$$(20) \quad v''''(\xi) - [q_1(\xi)v'(\xi)]' + q_2(\xi)v(\xi) = \lambda v(\xi),$$

where

$$\xi = \int_0^x \left[\frac{\rho(y)}{a(y)} \right]^{1/4} dy, \quad v(\xi) = \rho(x)^{3/8} a(x)^{1/8} u(x),$$

and $q_1(\xi)$, $q_2(\xi)$ are b -periodic expressions involving a and ρ (see [3]).

Then one can use in (20) the asymptotic estimates of [29], Part I, Chapt. II, to conclude that, in each sector

$$S_l = \left\{ \lambda \in \mathbb{C} : \frac{l\pi}{4} \leq \arg(\lambda) \leq \frac{(l+1)\pi}{4} \right\}, \quad l = 0, 1, \dots, 7,$$

of the complex λ -plane there are four λ -analytic linearly independent solutions $\phi_j(x; t; \lambda)$, $j = 1, 2, 3, 4$, of

$$(21) \quad [a(x; t)u''(x)]'' = \lambda \rho(x; t)u(x),$$

such that, given $M > 0$,

$$(22) \quad \left| \phi_j(x; t; \lambda) - \frac{e^{\varepsilon_j \lambda^{1/4} S(x; t)}}{\rho(x; t)^{3/8} a(x; t)^{1/8}} \right| \leq K \frac{|e^{\varepsilon_j \lambda^{1/4} S(x; t)}|}{|\lambda|^{1/4}}, \quad 0 \leq x \leq M,$$

$$(23) \quad \left| \phi_j'(x; t; \lambda) - \frac{\varepsilon_j \lambda^{1/4} e^{\varepsilon_j \lambda^{1/4} S(x; t)}}{\rho(x; t)^{1/8} a(x; t)^{3/8}} \right| \leq K |e^{\varepsilon_j \lambda^{1/4} S(x; t)}|, \quad 0 \leq x \leq M,$$

$$(24) \quad \left| \phi_j''(x; t; \lambda) - \frac{\rho(x; t)^{1/8}}{a(x; t)^{5/8}} \varepsilon_j^2 \lambda^{1/2} e^{\varepsilon_j \lambda^{1/4} S(x; t)} \right| \leq K |\lambda^{1/4} e^{\varepsilon_j \lambda^{1/4} S(x; t)}|, \quad 0 \leq x \leq M,$$

$$(25) \quad \left| \phi_j'''(x; t; \lambda) - \frac{\rho(x; t)^{3/8}}{a(x; t)^{7/8}} \varepsilon_j^3 \lambda^{3/4} e^{\varepsilon_j \lambda^{1/4} S(x; t)} \right| \leq K |\lambda^{1/2} e^{\varepsilon_j \lambda^{1/4} S(x; t)}|, \quad 0 \leq x \leq M,$$

where

$$S(x; t) = \int_0^x \left[\frac{\rho(y; t)}{a(y; t)} \right]^{1/4} dy$$

(in particular, $S(nb; t) = nb$, if $n \in \mathbb{Z}$). Here $\lambda^{1/4}$ stands for the principal branch of the fourth root (so that $\Re\{\lambda^{1/4}\} \geq 0$, $\Im\{\lambda^{1/4}\} \geq 0$), $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\} = \{i, -1, -i, 1\}$, and the (positive) constant K depends on $a(x)$, $\rho(x)$, and M , but not on t .

We finish this subsection with a useful lemma.

Lemma 1. *If $r_j(\lambda; t)$, $j = 1, 2, 3, 4$, are the Floquet multipliers of (21), where $t \in [0, 1]$, then*

$$(26) \quad \left| \frac{r_j(\lambda; t)}{e^{\varepsilon_j \lambda^{1/4} b}} - 1 \right| \leq \frac{K}{|\lambda|^{1/4}}, \quad j = 1, 2, 3, 4,$$

where $\lambda^{1/4}$ is the principal branch of the fourth root, $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\} = \{i, -1, -i, 1\}$, and the constant $K > 0$ depends on $a(x)$, $\rho(x)$, but not on t .

Proof. Let

$$T = [T_{jk}]_{1 \leq j, k \leq 4}$$

be the Floquet matrix of (21) with respect to the basis $\phi_j(x; t; \lambda)$, $j = 1, 2, 3, 4$, where $\phi_j(x; t; \lambda)$ is the solution of (21) that satisfies (22), (23), (24), and (25). It follows that, as $|\lambda| \rightarrow \infty$,

$$(27) \quad T_{jk} = e^{\varepsilon_k \lambda^{1/4} b} \left[\delta_{jk} + O\left(\lambda^{-1/4}\right) \right], \quad \text{uniformly in } t,$$

where δ_{jk} is the Kronecker delta. As we know, $r_j(\lambda; t)$, $j = 1, 2, 3, 4$, are the roots of the equation

$$r^4 - A(\lambda; t)r^3 + [B(\lambda; t) + 2]r^2 - A(\lambda; t)r + 1 = 0,$$

where

$$A(\lambda; t) = \sum_{j=1}^4 T_{jj}$$

and

$$B(\lambda; t) + 2 = \sum_{1 \leq j < k \leq 4} \begin{vmatrix} T_{jj} & T_{jk} \\ T_{kj} & T_{kk} \end{vmatrix}.$$

Thus (27) implies that, as $|\lambda| \rightarrow \infty$,

$$A(\lambda; t) = e^{\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right] + e^{-i\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right]$$

and

$$\begin{aligned} B(\lambda; t) + 2 = & e^{\lambda^{1/4} b} e^{-i\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right] + e^{\lambda^{1/4} b} e^{i\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right] \\ & + e^{-\lambda^{1/4} b} e^{-i\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right] + e^{-\lambda^{1/4} b} e^{i\lambda^{1/4} b} \left[1 + O\left(\lambda^{-1/4}\right) \right] \end{aligned}$$

uniformly in t . The statement of the lemma follows easily from these two formulas. \square

3.2. The theorems. The analysis that follows is inspired by the work [20] of W. Kohn for the second-order case (see also [1]). Consider again the function of formula (8), namely

$$F(\lambda; k) = B(\lambda) - 2A(\lambda) \cos(kb) + 4 \cos^2(kb).$$

This function is the analogue of the function

$$\Delta(\lambda) - 2 \cos(kb)$$

that appears in the analysis of the Hill operator (see, e.g., [1]). Notice that, although $F(\lambda; k)$ is usually viewed as an entire function of λ , it is entire in both λ and k . We want to generalize Theorems A and B for the case of complex k .

Theorem 1. *For a fixed $k \in \mathbb{C}$, let λ^* be a zero of $F(\lambda; k)$ of multiplicity $m(\lambda^*) > 1$. Then λ^* is also a zero of the entire function*

$$(28) \quad D(\lambda) = E'(\lambda)^2 - 4A'(\lambda)^2 E(\lambda).$$

Furthermore, $D(\lambda) \not\equiv 0$.

Proof. The derivative of $F(\lambda; k)$ with respect to λ satisfies

$$(29) \quad F_\lambda(\lambda^*; k) = B'(\lambda^*) - 2A'(\lambda^*) \cos(kb) = 0.$$

A straightforward calculation (using (16)) can verify that

$$(30) \quad A'(\lambda)^2 F(\lambda; k) - A(\lambda) A'(\lambda) F_\lambda(\lambda; k) + 2B'(\lambda) F_\lambda(\lambda; k) - F_\lambda(\lambda; k)^2 = \frac{D(\lambda)}{16},$$

where $D(\lambda)$ is given by (28). Thus, if $F(\lambda^*; k) = F_\lambda(\lambda^*; k) = 0$, then $D(\lambda^*) = 0$.

If $D(\lambda) \equiv 0$, then all zeros of $F(\lambda; k)$ should have been multiple. But, by Theorem A the zeros of $F(\lambda; k)$ are simple, if k is real and $0 < k < \pi/b$. Thus, $D(\lambda) \not\equiv 0$. \square

One important feature of Theorem 1 is that $D(\lambda)$ is independent of k . The next theorem is quite informative. To prove it we employ the technique described in the previous subsection.

Theorem 2. *All the zeros of the entire function $D(\lambda)$ of (28) are real and they are located as follows: (a) $D(\lambda)$ has exactly one (simple) zero in each gap of the spectrum $S(a, \rho)$ of (1) (with the understanding that, if the gap is closed, i.e., collapses to a double periodic or antiperiodic eigenvalue, say λ^* , then the simple zero of $D(\lambda)$ is λ^*); (b) $D(\lambda)$ has exactly two zeros (counting multiplicities) in each ψ -gap of the pseudospectrum of (1). In case (b), if the ψ -gap is open, then $D(\lambda)$ has exactly two simple zeros in it, whereas if the ψ -gap is closed, i.e., collapses to a point ν^* , where $\nu^* = \nu_{2n-1} = \nu_{2n}$, or $\nu^* = \nu'_{2n-1} = \nu'_{2n}$, for some $n = 1, 2, 3, \dots$ (see the statement of Theorem D), then ν^* is a double zero of $D(\lambda)$. There are no other zeros of $D(\lambda)$.*

Proof. We start by examining the zeros of $D(\lambda)$ in the ψ -gaps. Let (α, β) be a ψ -gap. If it is closed, i.e., if $\alpha = \beta$, then, by Theorem D we have $E(\alpha) = E'(\alpha) = 0$. Thus (28) implies that $D(\alpha) = D'(\alpha) = 0$ (i.e., α is a zero of $D(\lambda)$ of multiplicity ≥ 2). Next assume that $\alpha < \beta$. We then have $E(\alpha) = E(\beta) = 0$, while, by Remark 1, $E(\lambda) > 0$, for $\lambda \in (\alpha, \beta)$. It follows that there is a $\gamma \in (\alpha, \beta)$ such that $E'(\gamma) = 0$. Then (28) implies that

$$D(\alpha) = E'(\alpha)^2 > 0, \quad D(\beta) = E'(\beta)^2 > 0, \quad D(\gamma) = -4A'(\gamma)^2 E(\gamma) < 0.$$

Therefore, $D(\lambda)$ must have a zero in (α, γ) and one in (γ, β) . In conclusion, $D(\lambda)$ always has at least two zeros (counting multiplicities) in each ψ -gap, even if it is closed.

Next we consider the unperturbed case. Using $E_0(\lambda)$ of (17), namely

$$E_0(\lambda) = 4 \left[\cosh(\lambda^{1/4}b) - \cos(\lambda^{1/4}b) \right]^2,$$

and the fact that

$$A_0(\lambda) = 2 \left[\cosh(\lambda^{1/4}b) + \cos(\lambda^{1/4}b) \right],$$

we get from (28) that

$$(31) \quad D_0(\lambda) = \frac{16b^2}{\lambda^{3/2}} \sinh(\lambda^{1/4}b) \sin(\lambda^{1/4}b) \left[\cosh(\lambda^{1/4}b) - \cos(\lambda^{1/4}b) \right]^2$$

(in particular, $D_0(0) = 16b^8 \neq 0$). Hence the zeros of $D_0(\lambda)$ are exactly as the theorem describes them. As we deform the unperturbed case continuously, until we reach (1), $D(\lambda)$ will continue to have two zeros in each ψ -gap, unless some new nonreal zeros enter the ψ -gap. Also, by Theorem A and Theorem 1, the simple zero that $D(\lambda)$ has in each gap of the spectrum $S(a, \rho)$ (including the case of a closed gap) cannot move into the interior of a band. Hence this zero will remain in the gap until some other nonreal zero enters the gap. Therefore, as we deform the unperturbed case, $D(\lambda)$ will continue to have exactly one zero in each gap and exactly two zeros (counting multiplicities) in each ψ -gap, until some nonreal zero(s) enters a gap or a ψ -gap. But where can these nonreal zeros come from? Since $D_0(\lambda)$ has no other zeros, they can only come from infinity.

From the above it follows that, in order to finish the proof, we need to demonstrate that, as we deform the unperturbed case, no new zeros of $D(\lambda)$ can appear from infinity. Assume first that $a, \rho \in C^\infty(\mathbb{R})$. Let $D(\lambda^*) = 0$ where λ^* does not satisfy the statement of the theorem. Consider the continuous deformation (19) and let $D(\lambda; t)$ be the function $D(\lambda)$ for the problem (21). Then there is a $t_0 \in [0, 1)$ and a zero $\lambda_\omega(t)$ of $D(\lambda; t)$ such that

$$(32) \quad \lambda_\omega(1) = \lambda^* \quad \text{and} \quad \lim_{t \searrow t_0} |\lambda_\omega(t)| = \infty$$

($\lambda_\omega(t)$ depends continuously on t).

By Theorem 1 there is a $k = k(t) \in \mathbb{C}$ such that

$$(33) \quad F(\lambda_\omega(t); k; t) = 0$$

and

$$(34) \quad F_\lambda(\lambda_\omega(t); k; t) = 0.$$

An equivalent way to state (33) is to say that

$$r_j = r_j(\lambda_\omega(t); t) = e^{ikb}, \quad \text{for some } j = 1, 2, 3, 4,$$

namely

$$(35) \quad r_1 + r_4 = r_1 + r_1^{-1} = 2 \cos(kb) \quad \text{or} \quad r_2 + r_3 = r_2 + r_2^{-1} = 2 \cos(kb).$$

Using (29) and (35), formula (34) becomes

$$(36) \quad B_\lambda(\lambda_\omega(t); k; t) = A_\lambda(\lambda_\omega(t); k; t) \left[r_j(\lambda_\omega(t); t) + r_j(\lambda_\omega(t); t)^{-1} \right], \quad j = 1 \text{ or } 2.$$

By recalling (14), (36) becomes

$$\begin{aligned} & \left[r_2(\lambda_\omega) + r_2(\lambda_\omega)^{-1} \right] \partial_\lambda [r_1 + r_1^{-1}]|_{\lambda=\lambda_\omega} \\ & \quad + \left[r_1(\lambda_\omega) + r_1(\lambda_\omega)^{-1} \right] \partial_\lambda [r_2 + r_2^{-1}]|_{\lambda=\lambda_\omega} \\ & = \left[r_j(\lambda_\omega) + r_j(\lambda_\omega)^{-1} \right] \partial_\lambda [r_1 + r_1^{-1}]|_{\lambda=\lambda_\omega} \\ & \quad + \left[r_j(\lambda_\omega) + r_j(\lambda_\omega)^{-1} \right] \partial_\lambda [r_2 + r_2^{-1}]|_{\lambda=\lambda_\omega} \end{aligned}$$

where the dependence in t is suppressed for typographical convenience and $j = 1$ or 2 . Thus,

$$\begin{aligned} & \left[r_2(\lambda_\omega) + r_2(\lambda_\omega)^{-1} \right] \partial_\lambda [r_1 + r_1^{-1}]|_{\lambda=\lambda_\omega} \\ &= \left[r_1(\lambda_\omega) + r_1(\lambda_\omega)^{-1} \right] \partial_\lambda [r_1 + r_1^{-1}]|_{\lambda=\lambda_\omega}, \end{aligned}$$

if $j = 1$, or

$$\begin{aligned} & \left[r_1(\lambda_\omega) + r_1(\lambda_\omega)^{-1} \right] \partial_\lambda [r_2 + r_2^{-1}]|_{\lambda=\lambda_\omega} \\ &= \left[r_2(\lambda_\omega) + r_2(\lambda_\omega)^{-1} \right] \partial_\lambda [r_2 + r_2^{-1}]|_{\lambda=\lambda_\omega}, \end{aligned}$$

if $j = 2$. Therefore,

$$(37) \quad r_1(\lambda_\omega) + r_1(\lambda_\omega)^{-1} = r_2(\lambda_\omega) + r_2(\lambda_\omega)^{-1}$$

or

$$\left[1 - r_1(\lambda_\omega)^{-2} \right] \partial_\lambda r_1(\lambda_\omega) = 0$$

or

$$\left[1 - r_2(\lambda_\omega)^{-2} \right] \partial_\lambda r_2(\lambda_\omega) = 0.$$

An equivalent way to write the last two equations is

$$(38) \quad \left[r_1(\lambda_\omega)^2 - 1 \right] \left[r_2(\lambda_\omega)^2 - 1 \right] = 0$$

or

$$(39) \quad [\partial_\lambda r_1(\lambda_\omega)] [\partial_\lambda r_2(\lambda_\omega)] = 0.$$

Thus, if $D(\lambda_\omega; t) = 0$, then λ_ω satisfies (37), or (38), or (39). But (37) means that $E(\lambda_\omega; t) = 0$ and, by Theorem 1, this can happen if and only if λ_ω is a double zero of $E(\lambda; t)$. Similarly, (38) means that λ_ω is a periodic or antiperiodic eigenvalue of (21) and this can happen (see Theorem B) if and only if λ_ω is a double such eigenvalue, equivalently if λ_ω is a double zero of $F^+(\lambda; t)F^-(\lambda; t)$. Therefore, either $\lambda_\omega(t)$ is a double zero of

$$F^+(\lambda; t)F^-(\lambda; t)E(\lambda; t),$$

or $\lambda_\omega(t)$ satisfies (39).

Next, let $\{z_{-n}(t)\}_{n=0}^\infty$ be the zeros (counting multiplicities) of $E(\lambda; t)$ and $\{z_n(t)\}_{n=0}^\infty$ be the zeros (counting multiplicities) of $F^+(\lambda; t)F^-(\lambda; t)$ (thus $z_0(t) = 0$). Furthermore, we assume that $\{z_n(t)\}_{n=-\infty}^\infty$ is increasing in n . We then have the estimates (see [33])

$$(40) \quad |z_n(t) - z_n(0)| \leq Kn^2, \quad \text{for all } n \in \mathbb{Z},$$

where the (positive) constant K is independent of $t \in [0, 1]$, $z_0(0) = 0$, and, by (11) and (18),

$$(41) \quad z_{1-2l}(0) = z_{-2l}(0) = -4 \left(\frac{l\pi}{b} \right)^4, \quad z_{2l-1}(0) = z_{2l}(0) = \left(\frac{l\pi}{b} \right)^4, \quad l = 1, 2, 3, \dots$$

In particular, there is a constant $C > 0$ such that

$$(42) \quad z_{n+2}(0) - z_n(0) \geq C|n|^3, \quad \text{for all } n \in \mathbb{Z}.$$

Finally, recall that $z_n(t)$ is a branch point of $r(\lambda; t)$ if and only if $z_n(t)$ is a simple zero of $F^+(\lambda; t)F^-(\lambda; t)$ or $E(\lambda; t)$.

We will now show that (32) is impossible. This will imply that there is no λ^* , such that $D(\lambda^*) = 0$, violating the statement of the theorem. Let $\Gamma = \Gamma(\lambda_\omega(t), R)$, the circle in the complex plane with radius R , centered at $\lambda_\omega(t)$. We assume that R is small enough so that Γ does not enclose any branch point of $r(\lambda; t)$, i.e., any $(z_n(t))$ which is a simple zero of $F^+(\lambda; t)F^-(\lambda; t)$ or $E(\lambda; t)$. Then by Cauchy's integral formula we have

$$\partial_\lambda r_j(\lambda_\omega) - \frac{\varepsilon_j b}{4\lambda_\omega^{3/4}} e^{\varepsilon_j \lambda_\omega^{1/4} b} = \frac{1}{2\pi i} \int_\Gamma \frac{r_j(z) - e^{\varepsilon_j z^{1/4} b}}{(z - \lambda_\omega)^2} dz.$$

Thus, if $\lambda_\omega(t)$ satisfies (39),

(43)

$$\left| \frac{\varepsilon_j b}{4\lambda_\omega^{3/4}} e^{\varepsilon_j \lambda_\omega^{1/4} b} \right| \leq \frac{1}{2\pi R} \int_0^{2\pi} \left| r_j(z(\theta)) - e^{\varepsilon_j z(\theta)^{1/4} b} \right| d\theta, \quad z(\theta) = \lambda_\omega(t) + e^{i\theta} R.$$

If we assume, in accordance with (32), that

$$|\lambda_\omega(t)| \rightarrow \infty,$$

and take

$$R = |\lambda_\omega(t)|^{(1/2)+\varepsilon}, \quad \text{for some } \varepsilon \in (0, 1/4),$$

then (43) and (26) imply that there is a constant K independent of t such that

$$\left| \frac{1}{\lambda_\omega(t)^{3/4}} e^{\varepsilon_j \lambda_\omega^{1/4} b} \right| \leq \frac{K}{R} \frac{|e^{\varepsilon_j \lambda_\omega^{1/4} b}|}{|\lambda_\omega(t)|^{1/4}} = K \frac{|e^{\varepsilon_j \lambda_\omega^{1/4} b}|}{|\lambda_\omega(t)|^{(3/4)+\varepsilon}}.$$

This inequality is obviously impossible as $|\lambda_\omega(t)|$ gets arbitrarily large; hence Γ must enclose branch points. We have, therefore, established the following: Given $\varepsilon \in (0, 1/4)$, there is a constant Λ , independent of t , such that, if $|\lambda_\omega(t)| \geq \Lambda$, then

$$|\lambda_\omega(t) - z_n(t)| \leq |\lambda_\omega(t)|^{(1/2)+\varepsilon}, \quad \text{for some } n \in \mathbb{Z}.$$

Using (40) and (41) the above estimate can be written as

$$(44) \quad |\lambda_\omega(t) - z_n(0)| \leq K |n|^{2+4\varepsilon}, \quad \text{for some } n \in \mathbb{Z}.$$

Finally, since $\varepsilon \in (0, 1/4)$, the estimates (42) and (44) make it impossible for $\lambda_\omega(t)$ to move continuously to infinity. Hence (32) is impossible and the theorem is proved for $a, \rho \in C^\infty(\mathbb{R})$. The general case follows by approximation by C^∞ functions. We just have to observe that $D(\lambda)$ can be written in terms of the fundamental solutions $u_j(b; \lambda)$ (and their derivatives) and apply the proposition of the appendix. \square

Remark 2. If $D(\lambda^*) = 0$, then the theorem implies that λ^* is in a gap or a ψ -gap. Thus, the corresponding Floquet multipliers $r_j(\lambda^*)$, $j = 1, 2, 3, 4$, are all real (thus, $r_j(\lambda^*) = e^{ik_j}$ where $\Re\{k_j\} = 0$ or π/b). As we have already seen (see also Lemma 4 below), if λ^* is in a spectral gap, then we always have $r_1(\lambda^*), r_4(\lambda^*) > 0$, while $r_2(\lambda^*), r_3(\lambda^*)$ may be positive or negative; however, if λ^* is in a ψ -gap, then all $r_j(\lambda^*)$, $j = 1, 2, 3, 4$, have the same sign. Furthermore, if λ^* is in a gap or a ψ -gap, then (39) implies that $r'_j(\lambda^*) = 0$, for some $j = 1, 2, 3, 4$, and conversely.

The next theorem goes a little deeper. It completes Theorem 1.

Theorem 3. For a fixed $k \in \mathbb{C}$, the multiplicity m of any zero λ^* of $F(\lambda; k)$ can be either one or two (of course, by Theorem 1, $m = 2$ if and only if $D(\lambda^*) = 0$).

Proof. Let λ^* be a zero of $F(\lambda; k)$ with multiplicity $m \geq 3$. Then (30) implies that λ^* is a zero of $D(\lambda)$ of multiplicity at least $m - 1$. But, by Theorem 2, the multiplicity of any zeros of $D(\lambda)$ can be at most two. Thus, $m = 3$ and λ^* is a double zero of $D(\lambda)$. But then, again by Theorem 2, we must have that λ^* corresponds to a closed ψ -gap, i.e., it is a double zero of $E(\lambda)$; thus $E(\lambda^*) = E'(\lambda^*) = 0$. Next we observe that, using (16), we can write (30) as

$$A'(\lambda)^2 F(\lambda; k) - \frac{E'(\lambda)}{2} F_\lambda(\lambda; k) - F_\lambda(\lambda; k)^2 = \frac{D(\lambda)}{16},$$

which implies that λ^* is a zero of $D(\lambda)$ of multiplicity 3, a contradiction! Thus, $F(\lambda; k)$ cannot have any zeros of multiplicity larger than two. \square

We continue with a lemma which is by itself interesting since it characterizes the zeros of the entire functions $A(\lambda)$ and $B(\lambda)$.

Lemma 2. (a) *The set of zeros of $A(\lambda)$ of (4) is the $(\pi/(2b))$ -Floquet ψ -spectrum $\Psi_{\pi/2b}(a, \rho)$. Furthermore, all zeros of $A(\lambda)$ are simple.*

(b) *The set of zeros of $B(\lambda)$ of equality (9) is the $(\pi/(2b))$ -Floquet spectrum $\{\lambda_n(\pi/(2b))\}_{n=1}^\infty$. Again, all zeros of $B(\lambda)$ are simple.*

Proof. Part (a) follows immediately from (15) and Theorem C.

Part (b) also follows immediately from the fact (see (14)) that $B(\lambda) = (r_1 + r_1^{-1})(r_2 + r_2^{-1})$. \square

Remark 3. The lemma implies that $B(\lambda)$ has one zero (counting multiplicities) in the interior of each band of the spectrum of (1) and no other zeros. In particular, all zeros of $B(\lambda)$ are (real and) strictly positive and simple.

Likewise $A(\lambda)$ has one zero (counting multiplicities) in the interior of each ψ -band of the ψ -spectrum of (1) and no other zeros. In particular, all zeros of $A(\lambda)$ are strictly negative and simple.

From Theorems 1, 2, and 3, it follows that, if λ^* is a multiple zero of $F(\lambda; k)$, then $\lambda^* \in \mathbb{R}$, its multiplicity is two, and we must have $k \in \mathbb{C}$ with $\Re\{k\} = 0$ or π/b (without loss of generality). Using Lemma 1 we can, in addition, prove the following:

Theorem 4. *A given number λ^* can be a double zero of $F(\lambda; k)$ for at most one value of k with $\Re\{k\} \in \{0, \pi/b\}$.*

Proof. Assume

$$F(\lambda^*; k) = F_\lambda(\lambda^*; k) = 0.$$

Then (29) holds, namely

$$B'(\lambda^*) - 2A'(\lambda^*) \cos(kb) = 0.$$

If there are two distinct $k_1 \neq k_2$ with $\Re\{k_1\}, \Re\{k_2\} \in \{0, \pi/b\}$, for which the above is true, we must have

$$(45) \quad B'(\lambda^*) = A'(\lambda^*) = 0.$$

Now the functions $A(\lambda)$ and $B(\lambda)$ are entire of order $1/4$. Since the zeros of $A(\lambda)$ are positive and the zeros of $B(\lambda)$ are negative (see Remark 3 above), it follows by a well-known theorem of complex analysis (see, e.g., [37]) that the zeros of $A'(\lambda)$ are positive while the zeros of $B'(\lambda)$ are negative. Thus, (45) is impossible. \square

Remark 4. Let $D(\lambda^*) = 0$ where λ^* is in an **open** gap or ψ -gap. Then by Theorems 1, 2, 3, and 4 there is a unique $k \in \mathbb{C} \setminus \mathbb{R}$ with $\Re\{k\} = 0$ or π/b such that λ^* is a double zero of $F(\lambda; k)$. The four Floquet multipliers that correspond to λ^* are distinct (since λ^* lies in an open gap or ψ -gap) and one of them is $r = e^{ikb}$ (where $r \in \mathbb{R}$, $r \neq \pm 1$). Hence there is only one (up to linear independence, of course) Floquet solution $f(x)$ satisfying

$$f(x+b) = rf(x).$$

We can thus say that, for this particular k , the algebraic multiplicity of λ^* is two, but its geometric multiplicity is one!

If λ^* corresponds to a **closed** gap, i.e., λ^* is a double periodic or antiperiodic eigenvalue, then, by Theorem 2, $D(\lambda^*) = 0$ and hence λ^* is a double zero of $F(\lambda; k)$ for a unique (see Theorem 4) k that equals 0 in the periodic case, or equals π/b in the antiperiodic case. But now, by Theorem B, the geometric multiplicity of λ^* is also two.

Finally, if λ^* corresponds to a **closed** ψ -gap, then again, by Theorem 2, $D(\lambda^*) = 0$ and hence λ^* is a double zero of $F(\lambda; k)$ for a unique nonreal k with real part in $\{0, \pi/b\}$. In this case though (a kind of exception of the exception), as we will see later (Theorem 8) there are two linearly independent Floquet solutions with multiplier $r = e^{ikb}$; thus, the geometric multiplicity of λ^* is also two.

4. A CLOSER LOOK AT THE ENDPOINTS OF THE PSEUDOGAPS

4.1. The operators L_k . Let $k \in \mathbb{C}$ be given. As we have already seen, the eigenvalues of the problem

$$(46) \quad [a(x)u''(x)]'' = \lambda\rho(x)u(x), \quad u^{(j)}(b) = ru^{(j)}(0), \quad j = 0, 1, 2, 3,$$

where $r = e^{ikb}$, are the zeros of the function $F(\lambda; k)$ defined in (8).

Problems such as (46) (almost always related to operators with periodic coefficients) have an equivalent formulation (see, e.g., [21]):

Let L_k be the operator on $L^2_\rho(0, b)$ defined by

$$(47) \quad L_k v = \rho(x)^{-1} (d/dx + ik)^2 \left[a(x) (d/dx + ik)^2 v \right],$$

with periodic boundary conditions. Then $v(x)$ is an eigenfunction of L_k with corresponding eigenvalue λ if and only if $u(x) = e^{ikx}v(x)$ is an eigenfunction of (46) corresponding to the same eigenvalue λ (in other words, $u(x)$ is a Floquet solution of (1) with a prescribed multiplier $r = e^{ikb}$).

The adjoint operator of L_k is

$$L_k^* = L_{\bar{k}}.$$

In particular, L_k is selfadjoint if and only if $k \in \mathbb{R}$. Let $G_k(x, y; \lambda)$ be the Green function of L_k and $\tilde{G}_k(x, y; \lambda)$ be the Green function of (46). We have

$$\tilde{G}_k(x, y; \lambda) = e^{ik(x-y)} G_k(x, y; \lambda).$$

It is well known (see, e.g., [9]) that $G_k(x, y; \lambda)$ can be expressed as an expression which is entire in λ divided by $F(\lambda; k)$. In fact, $G_k(x, y; \lambda)$ is meromorphic in λ and its poles are the zeros of $F(\lambda; k)$. If λ^* is a double zero of $F(\lambda; k)$, then λ^* can be a simple or a double pole of $G_k(x, y; \lambda)$. In the latter case (see [9]), L_k and hence (46) may not possess a complete set of (proper) eigenfunctions. More precisely,

apart from the eigenfunction, say $\phi^*(x)$, that corresponds to λ^* , there may be a generalized eigenfunction $\psi^*(x)$:

$$(L_k - \lambda^*)\psi^*(x) = \phi^*(x).$$

In fact, it is known that if $a(x)\rho(x) \equiv 1$ (thus the problem is essentially of second order), then the above situation can actually happen (see [20]). We expect this anomaly to arise in the general case too, as long as λ^* is in an open gap or ψ -gap. These comments should be compared with Remark 4 above.

4.2. The endpoints of an open ψ -gap. The following theorem presents another case where an algebraic multiplicity is equal to the corresponding geometric. It can be viewed as a partial complement of Theorem B (in the sense that Theorem B is about endpoints of bands, while the theorem below is about endpoints of non-touching ψ -bands). The remaining case of touching ψ -bands (equivalently: closed ψ -gaps) is covered later by Theorem 8.

Theorem 5. *Let (α, β) be an open ψ -gap of (1) (i.e., $\alpha < \beta$) and $\nu = \alpha$ or $\nu = \beta$. Then the Floquet matrix $T(\nu)$ is similar to the matrix*

$$\begin{pmatrix} r_1 & 1 & 0 & 0 \\ 0 & r_1 & 0 & 0 \\ 0 & 0 & r_1^{-1} & 1 \\ 0 & 0 & 0 & r_1^{-1} \end{pmatrix}.$$

In other words, the equation

$$[a(x)u''(x)]'' = \nu\rho(x)u(x)$$

has exactly two linearly independent proper Floquet solutions, one with multiplier r_1 and one with r_1^{-1} .

Proof. Let $\nu = \alpha$ or β . Since (α, β) is a ψ -gap and $\alpha < \beta$, we have that $E(\nu) = 0$, but $E'(\nu) \neq 0$. Thus, by (28),

$$D(\nu) \neq 0.$$

In the complex λ -plane consider the open disk $B_\varepsilon(\nu)$, i.e., with center ν and radius ε . We choose $\varepsilon > 0$ small enough so that

$$D(\lambda) \neq 0 \quad \text{if } \lambda \in B_\varepsilon(\nu).$$

Let

$$r_1(\nu) = e^{ik(\nu)b} = r_2(\nu),$$

where $ik(\nu)$ or $ik(\nu) - (\pi/b)$ is real and, without loss of generality, strictly positive (if not, we consider r_j^{-1} instead of r_j). From (4), (9), and (16),

$$r_j(\lambda) + \frac{1}{r_j(\lambda)} = \frac{A(\lambda)}{2} \pm \frac{\sqrt{E(\lambda)}}{2}, \quad j = 1 \text{ or } 2,$$

where $\sqrt{\cdot}$ denotes the principal branch of the square root function. If $\lambda \in B_\varepsilon(\nu)$, then $E(\lambda) = E_1(\lambda)(\lambda - \nu)$, where $E_1(\lambda) \neq 0$ in $B_\varepsilon(\nu)$. Hence the above formula can be written as

$$r_1(\lambda) + \frac{1}{r_1(\lambda)} = \frac{A(\lambda)}{2} + \frac{\sqrt{E_1(\lambda)}}{2} \sqrt{\lambda - \nu}$$

and

$$(48) \quad r_2(\lambda) + \frac{1}{r_2(\lambda)} = \frac{A(\lambda)}{2} - \frac{\sqrt{E_1(\lambda)}}{2} \sqrt{\lambda - \nu}.$$

Assuming $|r_1(\lambda)|, |r_2(\lambda)| > 1$, the above equations give $r_1(\lambda) = e^{ik_1(\lambda)b}$ and $r_2(\lambda) = e^{ik_2(\lambda)b}$ uniquely. Of course, $k_1(\lambda) \neq k_2(\lambda)$ if $\lambda \in B_\varepsilon(\nu)$, $\lambda \neq \nu$.

If ε is sufficiently small, Theorem 1 implies that $F(\lambda; k(\nu))$ has exactly one zero in $B_\varepsilon(\nu)$, namely $\lambda = \nu$. In fact, there is a neighborhood N of $k(\nu)$ such that, if $k \in N$, then $F(\lambda; k)$ has exactly one zero (counting multiplicities) $\lambda = \lambda(k)$ in $B_\varepsilon(\nu)$. Furthermore, if $k \in N$, $k \neq k(\nu)$, then (1) has four proper Floquet solutions corresponding to this $\lambda = \lambda(k)$, with corresponding multipliers $r_1(\lambda) = e^{ikb}$, $r_1(\lambda)^{-1} = e^{-ikb}$, $r_2(\lambda)$, and $r_2(\lambda)^{-1}$, where $r_2(\lambda)$ is given by (48). Therefore, for $k \in N$, $k \neq k(\nu)$, we must have (since one of these four Floquet solutions corresponds to an eigenfunction of L_k)

$$\frac{1}{2\pi i} \int_{\partial B_\varepsilon(\nu)} \left[\int_0^b G_k(x, x; \lambda) dx \right] d\lambda = 1,$$

where, as in the previous subsection, $G_k(x, y; \lambda)$ is the Green function of L_k . Letting $k \rightarrow k(\nu)$ we obtain

$$(49) \quad \frac{1}{2\pi i} \int_{\partial B_\varepsilon(\nu)} \left[\int_0^b G_{k(\nu)}(x, x; \lambda) dx \right] d\lambda = 1,$$

which says that there is only one (proper) Floquet solution corresponding to $r_1(\nu) = r_2(\nu) = e^{ik(\nu)b}$ (if there were two Floquet solutions, the value of the integral in (48) would have been 2). Considering the adjoint case, which has an equivalent behavior (see [9], Ch. 12, Sec. 5), we can conclude that there is, also, only one Floquet solution corresponding to $r_1(\nu)^{-1} = r_2(\nu)^{-1} = e^{-ik(\nu)b}$. \square

Remark 5. If $E(\nu) = 0$, $\nu \neq 0$ (remember that $\nu \in \mathbb{R}$), then, since L_k and its adjoint L_k^* have the same number of proper and generalized eigenfunctions corresponding to ν (see [9], Ch. 12, Sec. 5), it follows that $T(\nu)$ is similar to one of the following matrices:

$$\begin{pmatrix} r_1 & 1 & 0 & 0 \\ 0 & r_1 & 0 & 0 \\ 0 & 0 & r_1^{-1} & 1 \\ 0 & 0 & 0 & r_1^{-1} \end{pmatrix}, \quad \begin{pmatrix} r_1 & 0 & 0 & 0 \\ 0 & r_1 & 0 & 0 \\ 0 & 0 & r_1^{-1} & 0 \\ 0 & 0 & 0 & r_1^{-1} \end{pmatrix}.$$

Theorem 5 covers the case where ν is a simple zero of $E(\lambda)$. The case where ν is a double zero of $E(\lambda)$, i.e., the corresponding ψ -gap is closed, is deeper, since in this case (49) becomes

$$\frac{1}{2\pi i} \int_{\partial B_\varepsilon(\nu)} \left[\int_0^b G_{k(\nu)}(x, x; \lambda) dx \right] d\lambda = 2.$$

Now, by Theorem 1, ν is a double zero of $F(\lambda; k(\nu))$. Consequently, the above equation says that either $L_{k(\nu)}$ has two proper linearly independent eigenfunctions and hence $T(\nu)$ is diagonalizable, or $L_{k(\nu)}$ has one proper and one generalized eigenfunction and hence $T(\nu)$ is not diagonalizable. Later, in Theorem 8, we will see that the latter can never happen, i.e., $T(\nu)$ is always diagonalizable.

5. A MULTIPOINT EIGENVALUE PROBLEM

In the Hill case, the Dirichlet spectrum $\{\mu_n\}_{n=1}^{\infty}$ (i.e., the eigenvalues corresponding to the boundary conditions $u(0) = u(b) = 0$) plays an important role in the general spectral theory, especially in the formulation and solution of the inverse spectral problem. We propose the following multipoint problem as an analogue of Hill's Dirichlet problem for the Euler-Bernoulli case:

$$(50) \quad [a(x)u''(x)]'' = \lambda \rho(x)u(x), \quad u(0) = u(b) = u(2b) = u(3b) = 0.$$

An eigenvalue of (50) is any value of λ for which (50) has a nontrivial solution. We call such a solution an eigenfunction of (50).

Physically the problem (50) describes the vibration of a (periodic) beam fixed at four points.

Let $u_j(x; \lambda)$, $j = 1, 2, 3, 4$, be the fundamental solutions of (1). Since every solution of (1) is a linear combination of the fundamental solutions, it follows that λ is an eigenvalue of (50) (that is, λ is such that (50) has a nontrivial solution) if and only if λ is a zero of the entire function

$$H(\lambda) \stackrel{\text{def}}{=} \begin{vmatrix} u_1(0; \lambda) & u_2(0; \lambda) & u_3(0; \lambda) & u_4(0; \lambda) \\ u_1(b; \lambda) & u_2(b; \lambda) & u_3(b; \lambda) & u_4(b; \lambda) \\ u_1(2b; \lambda) & u_2(2b; \lambda) & u_3(2b; \lambda) & u_4(2b; \lambda) \\ u_1(3b; \lambda) & u_2(3b; \lambda) & u_3(3b; \lambda) & u_4(3b; \lambda) \end{vmatrix}.$$

But $u_1(0; \lambda) = 1$ and $u_2(0; \lambda) = u_3(0; \lambda) = u_4(0; \lambda) = 0$; therefore,

$$(51) \quad H(\lambda) = \begin{vmatrix} u_2(b; \lambda) & u_3(b; \lambda) & u_4(b; \lambda) \\ u_2(2b; \lambda) & u_3(2b; \lambda) & u_4(2b; \lambda) \\ u_2(3b; \lambda) & u_3(3b; \lambda) & u_4(3b; \lambda) \end{vmatrix}.$$

We can, thus, say that the spectrum of (50) is the set of zeros of $H(\lambda)$. In particular,

$$H(0) = b^4 \left(\int_0^b \frac{dx}{a(x)} \right)^2;$$

thus 0 is not an eigenvalue of (50). In the unperturbed case we have

$$(52) \quad H_0(\lambda) = \frac{1}{\lambda^{3/2}} \sinh(\lambda^{1/4}b) \sin(\lambda^{1/4}b) \left[\cosh(\lambda^{1/4}b) - \cos(\lambda^{1/4}b) \right]^2.$$

Notice that by (31),

$$H_0(\lambda) = \frac{D_0(\lambda)}{16b^2}.$$

We continue with some properties of (50) and its accompanying function $H(\lambda)$. But first we need some lemmas.

Lemma 3. (a) *If there is a nontrivial function $u(x)$ such that*

$$\begin{aligned} [a(x)u''(x)]'' &= \lambda \rho(x)u(x), & b_1 < x < b_2, \\ u(b_1) &= u'(b_1) = u(b_2) = u'(b_2) = 0, \end{aligned}$$

then λ is real and strictly positive.

(b) Likewise, if there is a nontrivial function $u(x)$ such that

$$\begin{aligned} [a(x)u''(x)]'' &= \lambda \rho(x)u(x), & x \in (b, \infty) \quad (\text{or } x \in (-\infty, b)), \\ u(b) = u'(b) &= 0, & u \in L^2(b, \infty) \quad (\text{resp. } u \in L^2(-\infty, b)), \end{aligned}$$

then λ is real and strictly positive.

Proof. (a) Multiplying the equation by $\overline{u(x)}$ and integrating yields

$$\int_{b_1}^{b_2} [a(x)u''(x)]'' \overline{u(x)} dx = \lambda \int_{b_1}^{b_2} \rho(x)u(x)\overline{u(x)} dx.$$

Next, by applying integration by parts (twice) in the left-hand side and using the boundary conditions, we obtain

$$\int_{b_1}^{b_2} a(x)u''(x)\overline{u''(x)} dx = \lambda \int_{b_1}^{b_2} \rho(x)u(x)\overline{u(x)} dx.$$

The assumption that $u(x)$ is not trivial, together with the boundary conditions $u(b_1) = u'(b_1) = 0$, imply that $u(x)$ is not a linear function. Thus both integrals in the above formula are strictly positive (remember $a(x), \rho(x) > 0$, for all x); hence $\lambda > 0$.

(b) The proof of this part is very similar to the proof of part (a). Assume that $u \in L^2(b, \infty)$. Since $u(x)$ is a linear combination of Floquet solutions (possibly including generalized ones), it follows that $u(x)$ and its derivatives decay exponentially, as $x \rightarrow \infty$; thus, we can apply again integration by parts and get

$$\int_b^\infty a(x)u''(x)\overline{u''(x)} dx = \lambda \int_b^\infty \rho(x)u(x)\overline{u(x)} dx;$$

hence, again $\lambda > 0$. □

The following lemma is contained in [31]. We include it here for the sake of completeness.

Lemma 4. *Let $\lambda > 0$. If the Floquet multipliers are indexed so that $|r_1| \geq |r_2| \geq |r_3| \geq |r_4|$, then ($r_3 = r_2^{-1}$ and)*

$$(53) \quad r_1 > |r_2| \geq |r_3| > r_4 = r_1^{-1}.$$

Furthermore, the Floquet solutions $f_1(x)$ and $f_4(x)$ corresponding to r_1 and r_4 never vanish.

Proof. If $\lambda > 0$, $u_1(x; \lambda)$, the first fundamental solution of (1), and $u_1(-x; \lambda)$ are increasing when $x \geq 0$. They actually grow exponentially. If λ is in the spectrum of (1), $|r_2| = |r_3| = 1$ and thus (53) and the statement about $f_1(x)$ and $f_4(x)$ must be true (remember $f_j(x) = e^{w_j x} p_j(x)$, where $p_j(x)$ is b -periodic and $r_j = e^{w_j b}$); otherwise, there would not be any exponentially growing solutions. Similarly, if λ is not in the spectrum, r_2 and $r_3 = r_2^{-1}$ are real. If we take the period of (1) to be $2b$, then the Floquet multipliers become

$$r_1^2 \geq r_2^2 > r_3^2 \geq r_4^2.$$

But the above inequalities become equalities only if λ is a zero of $E(\lambda)$. Since $\lambda > 0$ and the zeros of $E(\lambda)$ are nonpositive, we must have

$$|r_1| > |r_2| > |r_3| > |r_4|.$$

Now

$$u_1(x) = c_1 f_1(x) + c_2 f_2(x) + c_3 f_3(x) + c_4 f_4(x).$$

Hence the exponential growth of $u_1(x)$ and $u_1(-x)$ implies that r_1 and r_4 are positive and $f_1(x)$ and $f_4(x)$ do not change sign. \square

The next theorem should be compared with the property of the Hill operator stating that the Dirichlet eigenvalues are simple and their corresponding eigenfunctions are Floquet solutions [39]. The case left open (namely when μ is also a simple periodic or antiperiodic eigenvalue) is covered later by Theorem 7.

Theorem 6. *Let μ be an eigenvalue of (50). If $V(\mu)$ denotes the corresponding eigenspace, namely the vector space of all eigenfunctions of (50) associated to μ , then $\dim V(\mu) = 1$ or 2 . Furthermore, $V(\mu)$ always contains a proper Floquet solution; if $\dim V(\mu) = 2$, then $V(\mu)$ always contains two linearly independent proper Floquet solutions, except possibly in the case where μ is also a simple periodic or antiperiodic eigenvalue of (1) (in fact, we will see later, in Theorem 7, that this exception can never happen).*

Proof. Let $H(\mu) = 0$. Assume that $\dim V(\mu) = 3$ (clearly $\dim V(\mu) < 4$). Then we have three linearly independent eigenfunctions $\phi_1(x)$, $\phi_2(x)$, and $\phi_3(x)$ corresponding to μ . Let

$$\phi(x) = c_1 \phi_1(x) + c_2 \phi_2(x) + c_3 \phi_3(x).$$

We have $\phi(0) = \phi(b) = 0$. Also, we can choose c_1 , c_2 , and c_3 (not all zero) so that $\phi'(0) = \phi'(b) = 0$. Hence, Lemma 3 implies that $\mu > 0$. But then, Lemma 4 implies that the associated (to μ) Floquet solutions $f_1(x)$ and $f_4(x)$ never vanish. Furthermore, the space spanned by $\phi_1(x)$, $\phi_2(x)$, and $\phi_3(x)$ and the space spanned by $f_1(x)$ and $f_4(x)$ must have a nontrivial intersection. Thus there is an eigenfunction of (50) of the form

$$\gamma_1 f_1(x) + \gamma_4 f_4(x).$$

But this implies easily that

$$f_1(0; \mu) f_4(0; \mu) = 0,$$

a contradiction. Thus $\dim V(\mu) < 3$.

We now prove the rest of the theorem (see Theorem B and formula (16) for the definition of F^+ , F^- , and E that appear below).

Case $F^+(\mu)F^-(\mu)E(\mu) \neq 0$. Then, for $\lambda = \mu$, (1) possesses four distinct Floquet multipliers $r_j = r_j(\mu)$, $j = 1, 2, 3, 4$; therefore, it has four linearly independent (proper) Floquet solutions $f_j(x) = f_j(x; \mu)$, $j = 1, 2, 3, 4$, with

$$f_j(x+b) = r_j f_j(x).$$

If, in addition, μ is in the spectrum of (50), then

$$\begin{vmatrix} f_1(0) & f_2(0) & f_3(0) & f_4(0) \\ f_1(b) & f_2(b) & f_3(b) & f_4(b) \\ f_1(2b) & f_2(2b) & f_3(2b) & f_4(2b) \\ f_1(3b) & f_2(3b) & f_3(3b) & f_4(3b) \end{vmatrix} = 0;$$

namely,

$$f_1(0; \mu) f_2(0; \mu) f_3(0; \mu) f_4(0; \mu) \begin{vmatrix} 1 & 1 & 1 & 1 \\ r_1 & r_2 & r_3 & r_4 \\ r_1^2 & r_2^2 & r_3^2 & r_4^2 \\ r_1^3 & r_2^3 & r_3^3 & r_4^3 \end{vmatrix} = 0$$

or

$$f_1(0; \mu) f_2(0; \mu) f_3(0; \mu) f_4(0; \mu) \prod_{1 \leq j < l \leq 4} (r_l - r_j) = 0.$$

But the r_j 's are distinct; thus,

$$(54) \quad f_1(0; \mu) f_2(0; \mu) f_3(0; \mu) f_4(0; \mu) = 0.$$

This means that some Floquet solution, say $f_1(x; \mu)$, is an eigenfunction of (50).

Next, let $\phi(x)$ be another eigenfunction of (50) corresponding to μ . That is, $\phi(x)$ and $f_1(x)$ are linearly independent. It follows that there is a constant c_1 such that $\tilde{\phi}(x) = \phi(x) - c_1 f_1(x)$ is an eigenfunction of (50) and

$$\tilde{\phi}(x) = c_2 f_2(x) + c_3 f_3(x) + c_4 f_4(x).$$

But this implies easily that

$$f_2(0; \mu) f_3(0; \mu) f_4(0; \mu) = 0,$$

which, again, means that some Floquet solution f_j , $j = 2, 3, 4$, say $f_2(x; \mu)$, is an eigenfunction of (50).

Case $E(\mu) = 0$. That means that $\mu < 0$ and $r_1 = r_2 = r_3^{-1} = r_4^{-1}$. If we have coexistence of two Floquet solutions $f_1(x)$ and $f_2(x)$ with multiplier r_1 , then (see Remark 5) we also have coexistence of two Floquet solutions, $f_3(x)$ and $f_4(x)$, with multiplier r_1^{-1} . We can then find constants c_1 , c_2 , c_3 , and c_4 , such that $c_1 f_1(x) + c_2 f_2(x)$ and $c_3 f_3(x) + c_4 f_4(x)$ are in $V(\mu)$. If we do not have coexistence, then (see Remark 5) we have two proper Floquet solutions $f_1(x)$ and $f_3(x)$ with corresponding multipliers r_1 and r_1^{-1} , and two generalized Floquet solutions $g_1(x)$ and $g_3(x)$ satisfying

$$(55) \quad g_1(x+b) = r_1 g_1(x) + d_1 f_1(x), \quad g_3(x+b) = r_1^{-1} g_3(x) + d_3 f_3(x), \quad d_1 d_3 \neq 0.$$

Since $H(\mu) = 0$, we must have

$$\begin{vmatrix} f_1(0) & g_1(0) & f_3(0) & g_3(0) \\ f_1(b) & g_1(b) & f_3(b) & g_3(b) \\ f_1(2b) & g_1(2b) & f_3(2b) & g_3(2b) \\ f_1(3b) & g_1(3b) & f_3(3b) & g_3(3b) \end{vmatrix} = 0,$$

which implies

$$f_1(0; \mu)^2 f_3(0; \mu)^2 = 0.$$

The last equality says that one Floquet solution, say $f_1(x)$, is also an eigenfunction of (50). If $\phi(x)$ is another eigenfunction of (50) corresponding to μ , then there is an eigenfunction of the form

$$\tilde{\phi}(x) = c_1 g_1(x) + c_3 f_3(x) + c_4 g_3(x).$$

This implies that $g_1(x) \in V(\mu)$ or $f_3(x) \in V(\mu)$.

From (50) and the fact that $f_1(x)$ has the form

$$(56) \quad f_1(x) = e^{w_1 x} p_1(x) \quad \text{with} \quad r_1 = e^{w_1 b} \quad \text{and} \quad p_1(x+b) = p_1(x),$$

it follows that $g_1(x)$ has the form

(57)

$$g_1(x) = [p_2(x) + \beta p_1(x)x] e^{w_1 x} \quad \text{with} \quad \beta = \frac{d_1}{r_1 b} \quad \text{and} \quad p_2(x+b) = p_2(x).$$

Let us assume $g_1(x) \in V(\mu)$. Since we also have $g_1(x) \in V(\mu)$, it follows that $p_1(0) = p_2(0) = 0$; therefore,

$$f_1(nb) = g_1(nb) = 0, \quad \text{for all } n \in \mathbb{Z}.$$

Next notice that

$$f_1'(nb) = r_1^n p_1'(0) \neq 0 \quad \text{and} \quad g_1'(nb) = r_1^n [nb p_1'(0) + p_2'(0)] \quad \text{for all } n \in \mathbb{Z},$$

where $p_1'(0) \neq 0$ follows from Lemma 3. Introduce

$$(58) \quad v(x) = p_2'(0) f_1(x) - p_1'(0) g_1(x),$$

so that

$$(59) \quad [a(x)v''(x)]'' = \mu \rho(x)v(x)$$

and (for all $n \in \mathbb{Z}$)

$$(60) \quad v(nb) = 0, \quad v'(0) = 0.$$

By (59),

$$\int_0^{nb} [a(x)v''(x)]'' \overline{v(x)} dx = \mu \int_0^{nb} \rho(x)v(x)\overline{v(x)} dx.$$

Then, integration by parts and (60) yield (recall that $a(nb) = a(0)$)

$$(61) \quad -a(0)v''(nb)\overline{v'(nb)} + \int_0^{nb} a(x)v''(x)\overline{v''(x)} dx = \mu \int_0^{nb} \rho(x)v(x)\overline{v(x)} dx.$$

If we set

$$\varepsilon = \varepsilon(r_1) = \begin{cases} -1, & \text{if } |r_1| > 1, \\ 1, & \text{if } |r_1| < 1, \end{cases}$$

then (56) and (57) imply that $f_1(x)$, $g_1(x)$, and their derivatives decay exponentially, as $x \rightarrow \varepsilon\infty$, and by (58) the same is true for $v(x)$. Thus (61) implies that

$$\int_0^{\varepsilon\infty} a(x)v''(x)\overline{v''(x)} dx = \mu \int_0^{\varepsilon\infty} \rho(x)v(x)\overline{v(x)} dx;$$

in particular, $\mu > 0$, a contradiction. Therefore, $g_1(x) \notin V(\mu)$ and we are left with the only alternative, namely that $f_3(x) \in V(\mu)$.

Case $F^+(\mu)F^-(\mu) = 0$. That means that $\mu > 0$ is a periodic or antiperiodic eigenvalue. If we have coexistence of two periodic or antiperiodic Floquet solutions, then a linear combination of these can produce an eigenfunction of (50), and there is no other eigenfunction, i.e., $\dim V(\mu) = 1$ (since, by Lemma 4, the other two Floquet solutions never vanish). If there is only one periodic (or antiperiodic) Floquet solution, say $f_2(x)$, then there is a generalized Floquet solution $g_2(x)$, satisfying

$$g_2(x+b) = \varepsilon g_2(x) + c_2 f_2(x), \quad c_2 \neq 0,$$

where $\varepsilon = 1$, if $F^+(\mu) = 0$, and $\varepsilon = -1$, if $F^-(\mu) = 0$. In this case, $H(\mu) = 0$ implies that

$$\begin{vmatrix} f_1(0) & f_2(0) & g_2(0) & f_4(0) \\ f_1(b) & f_2(b) & g_2(b) & f_4(b) \\ f_1(2b) & f_2(2b) & g_2(2b) & f_4(2b) \\ f_1(3b) & f_2(3b) & g_2(3b) & f_4(3b) \end{vmatrix} = 0,$$

which implies easily that

$$f_1(0; \mu) f_2(0; \mu)^2 f_4(0; \mu) = 0;$$

hence there is a Floquet solution in $V(\mu)$. But, if $f_2(x) \in V(\mu)$ and $\dim V(\mu) = 2$ (this is, however, impossible, as we will see later in Theorem 7), we cannot, for the moment, exclude the possibility that $g_2(x) \in V(\mu)$. \square

Remark 6. One part of Theorem 6 states that the geometric multiplicity $m_g(\mu)$ of any eigenvalue μ of (50) cannot exceed two. We can define the algebraic multiplicity $m_a(\mu)$ of μ to be its multiplicity as a zero of $H(\lambda)$. From the above proof it follows easily that $m_a(\mu) \geq m_g(\mu)$. Theorems 7 and 8 below establish the equality of the two multiplicities.

The lemma that follows is needed for the proof of Theorem 7 below.

Lemma 5. *There are no zeros of $H(\lambda)$ in the interior of the bands or in the interior of the ψ -bands.*

Proof. Assume that μ is in the interior of a band. Then $F^+(\mu)F^-(\mu)E(\mu) \neq 0$; hence (54) must hold. By Lemma 4 we have that $f_1(x; \mu)$ and $f_4(x; \mu)$ never vanish,

$$r_1(\mu) = r_4(\mu)^{-1} > 1 \quad \text{and} \quad |r_2(\mu)| = |r_3(\mu)| = 1,$$

where $r_3(\mu) = r_2^{-1}(\mu) = \overline{r_2(\mu)} \notin \mathbb{R}$. Also, $f_3(x; \mu) = \overline{f_2(x; \mu)}$. Therefore,

$$(62) \quad f_2(0; \mu) = f_3(0; \mu) = 0.$$

We can, thus, write a linear combination

$$f(x) = d_2 f_2(x; \mu) + d_3 f_3(x; \mu)$$

such that $f(0) = f'(0) = 0$. This means that $f(x)$ can be written as a linear combination of the fundamental solutions $u_3(x; \mu)$ and $u_4(x; \mu)$, namely

$$(63) \quad f(x) = \gamma_3 u_3(x; \mu) + \gamma_4 u_4(x; \mu).$$

But (see [31]) $u_3(x; \mu) \rightarrow \infty$, as $x \rightarrow \pm\infty$, while $u_4(x; \mu) \rightarrow \pm\infty$, as $x \rightarrow \pm\infty$. Since $f(x)$ is bounded we can conclude that (63) is impossible. Thus (62) is impossible and $H(\mu) \neq 0$.

Assume now that μ is in the interior of a ψ -band (hence $\mu < 0$). Then there is a $k \in (0, \pi/b)$ such that

$$r_1(\mu) = \frac{1}{r_4(\mu)} = \overline{r_2(\mu)} = \frac{1}{r_3(\mu)} = |r_1(\mu)| e^{ikb}.$$

If $H(\mu) = 0$, then (54) implies that $f_j(0; \mu)$, for some j . Let us assume that

$$f_1(0; \mu) = 0.$$

We have

$$f_1(x; \mu) = e^{\alpha x} e^{ikx} p_1(x) \quad \text{and} \quad f_2(x; \mu) = \overline{f_1(x; \mu)} = e^{\alpha x} e^{-ikx} \overline{p_1(x)},$$

where $\alpha \in R \setminus \{0\}$ and $p_1(x+b) = p_1(x)$. Thus

$$f_2(0; \mu) = 0$$

and $f_1(\cdot; \mu), f_2(\cdot; \mu) \in L^2(0, \infty)$ (if $\alpha < 0$) or $f_1(\cdot; \mu), f_2(\cdot; \mu) \in L^2(-\infty, 0)$ (if $\alpha > 0$). We can, therefore, write a nontrivial linear combination

$$u(x) = c_1 f_1(x; \mu) + c_1 f_1(x; \mu)$$

that satisfies all assumptions of part (b) of Lemma 3 (with $b = 0$). Hence $\mu > 0$, a contradiction. Therefore, $H(\mu) \neq 0$. \square

We are now ready to prove our main theorem regarding the spectrum of (51), i.e., the zeros of $H(\lambda)$. The statement of the theorem resembles the one of Theorem 2.

Theorem 7. *All zeros of $H(\lambda)$, of (51), are real and they are located as follows: (a) $H(\lambda)$ has exactly one (simple) zero in the closure of each gap of the spectrum $S(a, \rho)$ (with the understanding that, if the gap is closed, i.e., collapses to a double periodic or antiperiodic eigenvalue, say λ^* , then the simple zero of $H(\lambda)$ is λ^*); (b) $H(\lambda)$ has exactly two zeros (counting multiplicities) in the closure of each ψ -gap of the pseudospectrum. In case (b), if the ψ -gap is closed, i.e., collapses to a point ν^* , where $\nu^* = \nu_{2n-1} = \nu_{2n}$, or $\nu^* = \nu'_{2n-1} = \nu'_{2n}$, for some $n = 1, 2, 3, \dots$ (see the statement of Theorem D), then ν^* is a double zero of $H(\lambda)$. There are no other zeros of $H(\lambda)$.*

Proof. As in the case of Theorem 2 (see also the appendix), we only need to prove the theorem for the case of smooth $a(x)$ and $\rho(x)$. Hence, from now on we assume $a, \rho \in C^\infty(\mathbb{R})$.

By (52) the theorem is valid in the unperturbed case $a(x) \equiv \rho(x) \equiv 1$. For general $a(x)$ and $\rho(x)$ let us consider the deformation (19) and the solutions $\phi_j(x; \lambda; t)$, $j = 1, 2, 3, 4$, satisfying (22). We set

$$(64) \quad \tilde{H}(\lambda; t) = \begin{vmatrix} \phi_1(0; t) & \phi_2(0; t) & \phi_3(0; t) & \phi_4(0; t) \\ \phi_1(b; t) & \phi_2(b; t) & \phi_3(b; t) & \phi_4(b; t) \\ \phi_1(2b; t) & \phi_2(2b; t) & \phi_3(2b; t) & \phi_4(2b; t) \\ \phi_1(3b; t) & \phi_2(3b; t) & \phi_3(3b; t) & \phi_4(3b; t) \end{vmatrix},$$

where the dependence of ϕ_j in λ has been suppressed for typographical convenience. For small $|\lambda|$'s, the ϕ_j 's can be chosen so that $\tilde{H}(0; t)$ stays away from 0, uniformly in t . Notice that $\tilde{H}(\lambda; t)$ and $H(\lambda; t)$ have the same zeros. The multiplicities of their zeros also agree (this can be easily checked when $t = 0$; then, since t is moving continuously, the multiplicities of two corresponding (i.e., equal) zeros, one of H and one of \tilde{H} , cannot suddenly become different).

In order to use the technique described in Subsection 3.1, we first need to estimate the large-magnitude zeros of $\tilde{H}(\lambda; t)$. First, consider $\{\mu_n(0)\}_{n \in \mathbb{Z}}$, the set of zeros of $\tilde{H}(\lambda; 0)$, counting multiplicities. As we have seen, $\tilde{H}(\lambda; 0)$ and $D_0(\lambda)$ have the same zeros; thus, (35) gives

$$\mu_{1-2l}(0) = \mu_{-2l}(0) = -4 \left(\frac{l\pi}{b} \right)^4, \quad \mu_{2l-1}(0) = \mu_{2l}(0) = \left(\frac{l\pi}{b} \right)^4, \quad l = 1, 2, 3, \dots$$

If μ is a large-magnitude zero of $\tilde{H}(\lambda; t)$, then (22) and (64) imply that

$$e^{\varepsilon_j \mu^{1/4} b} \left[1 + O\left(\frac{1}{\mu^{1/4}}\right) \right] = e^{\varepsilon_l \mu^{1/4} b} \left[1 + O\left(\frac{1}{\mu^{1/4}}\right) \right]$$

(uniformly in t), where $\varepsilon_j, \varepsilon_l \in \{i, -1, -i, 1\}$, $\varepsilon_j \neq \varepsilon_l$. This, in turn, implies that there is a $K > 0$ (independent of t) such that

$$\left| \mu^{1/4} - \mu_n(0)^{1/4} \right| \leq \frac{K}{|\mu|^{1/4}}, \quad \text{for some } n \in \mathbb{Z},$$

from which it follows that

$$|\mu - \mu_n(0)| \leq K n^2, \quad \text{for some } n \in \mathbb{Z}.$$

In other words, if $H(\mu; t) = 0$ and $|\mu|$ is sufficiently large, then there is an integer n such that μ is within distance $K n^2$ from $\mu_n(0)$. On the other hand, as we have already seen in (42), there is a constant $C > 0$ such that

$$\mu_{n+2}(0) - \mu_n(0) \geq C |n|^3, \quad \text{for all } n \in \mathbb{Z}.$$

Therefore, as in the proof of Theorem 2, no new zeros of $H(\lambda; t)$ can come from infinity, as we move t .

Thus, by the above discussion and Lemma 5, the only way in which the theorem can be violated is if some zeros of $H(\lambda; t)$ become nonreal. As we start moving t , the zeros in the gaps (including closed gaps) are all simple. Thus, the zeros that can first leave the real axis (in pairs of complex conjugates, of course) are the zeros in the ψ -gaps.

Let (α, β) be a ψ -gap (α, β vary continuously with t). For λ in $[\alpha, \beta]$, let $r_1, r_2, r_3 = r_2^{-1}$, and $r_4 = r_1^{-1}$, with $|r_1| \geq |r_2| > 1$, be the corresponding Floquet multipliers of

$$[a(x; t)u''(x)]'' = \lambda \rho(x; t)u(x).$$

Since (α, β) is a ψ -gap, there is a $\delta > 0$, independent of t , such that

$$|r_1| \geq |r_2| \geq 1 + \delta.$$

Now let $D \subset \mathbb{C}$ be a domain (depending on t) such that $[\alpha, \beta] \subset D$. If D is sufficiently small, then, as functions of λ , $f_j(x; \lambda; t)$ and, in particular, $f_j(0; \lambda; t)$, $j = 1, 2, 3, 4$, are analytic in D , with the only singularities being the branch points α, β (if $\alpha \neq \beta$). As λ moves around one of these branch points, $f_1(x; \lambda)$ becomes $f_2(x; \lambda)$ and $f_4(x; \lambda)$ becomes $f_3(x; \lambda)$.

Initially (i.e., when $t = 0$), $\mu = \alpha = \beta$ is a double zero of $H(\lambda; 0)$ and there are two (linearly independent) Floquet solutions $f_1(x; \mu)$ and $f_4(x; \mu)$, with multipliers r_1 and r_4 respectively, satisfying $f_1(0; \mu) = f_4(0; \mu) = 0$ ($f_1(x; \mu)$ and $f_4(x; \mu)$ are eigenfunctions of (50) corresponding to the eigenvalue μ). As we move t , $f_1(0; \lambda)$ (or $f_2(0; \lambda)$, if we encounter a branch point) will continue to have a zero in $[\alpha, \beta]$, and so will $f_4(0; \lambda)$ (or $f_3(0; \lambda)$, if we encounter a branch point). In order for a zero of $H(\lambda; t)$ to escape from the real axis, it first has to become double. So let us assume that, for $t = t_0$, $\mu = \mu(t_0) \in \mathbb{R}$ satisfies $f_j(0; \mu(t_0)) = f_l(0; \mu(t_0)) = 0$, $j \neq l$, i.e., $\mu(t_0)$ is a double zero of $H(\lambda; t_0)$. Notice that $f_j(x)$ and $f_l(x)$ must belong, one to $L^2(-\infty, 0)$ and one to $L^2(0, \infty)$. This follows by continuity, but, also, by Lemma 3 (if both $f_j(x)$ and $f_l(x)$ are, say, in $L^2(0, \infty)$, then we can construct a nontrivial solution $v(x) = c_j f_j(x) + c_l f_l(x)$, such that $v(0) = v'(0) = 0$ and $v \in L^2(0, \infty)$, which implies that $\mu(t_0) > 0$, a contradiction). Thus, without loss

of generality, we can take $j = 1$ and $l = 4$. Assume that, as t gets larger than t_0 , then immediately $\mu(t_0)$ splits into two nonreal zeros $\mu(t)$ and $\overline{\mu(t)}$ of $H(\lambda; t)$. The corresponding Floquet solutions are $f_1(x; \mu(t))$ (or $f_2(x; \mu(t))$, if $\mu(t_0)$ is a branch point) and $f_4(x; \overline{\mu(t)})$ (or $f_3(x; \overline{\mu(t)})$, if $\mu(t_0)$ is a branch point). They also have to be complex conjugates. But this is a contradiction, since, on the one hand, $f_1(x)$, $f_2(x)$ are in $L^2(-\infty, 0)$ and $f_3(x)$, $f_4(x)$ are in $L^2(0, \infty)$; but, on the other hand, a function and its complex conjugate are in the same L^2 -space. Therefore, the zeros of $H(\lambda; t)$ can never leave the real axis and the theorem is proved. \square

Remark 7. Since, by the above theorem, all the positive zeros of $H(\lambda)$ are simple, the case in Theorem 6 that was left unanswered, namely when μ is also a simple periodic or antiperiodic eigenvalue, is now decided: We always have $\dim V(\mu) = 1$.

Remark 8. If the coefficients of (1) are even functions, namely if

$$a(-x) = a(x) \quad \text{and} \quad \rho(-x) = \rho(x),$$

then for every solution $u(x)$ of (1) we have that $v(x) = u(-x)$ is also a solution. In particular, if $f(x) = e^{wx}p(x)$ is a Floquet solution, so is $f(-x) = e^{-wx}p(-x)$. It follows that, in this case, the positive zeros of $H(\lambda)$, i.e., the positive eigenvalues of (50), must also be periodic or antiperiodic eigenvalues, while the negative zeros of $H(\lambda)$ are all double, since, if $\mu < 0$ is such that $H(\mu) = 0$, then there are two eigenfunctions of (50), $f(x) = e^{wx}p(x)$ and $f(-x) = e^{-wx}p(-x)$ (one in $L^2(-\infty, 0)$ and the other in $L^2(0, \infty)$).

Remark 9. For $\xi \in \mathbb{R}$, consider the one-parameter family of shifted or translated functions $a_\xi(x)$ and $\rho_\xi(x)$ of $a(x)$ and $\rho(x)$, namely

$$(65) \quad a_\xi(x) = a(x + \xi) \quad \text{and} \quad \rho_\xi(x) = \rho(x + \xi),$$

and let $T(\lambda; \xi)$ be the corresponding Floquet matrices. Then $T(\lambda; \xi)$ is similar to $T(\lambda)$, for all ξ . In particular, this one-parameter family is isospectral and “iso-pseudospectral”. On the other hand, the spectrum $\{\mu_n(\xi)\}_{n \in \mathbb{Z}}$ of (50) evolves with ξ (thus we have an isospectral and iso-pseudospectral flow). It will be interesting to understand the evolution of $\mu_n(\xi)$ ’s with ξ , since this might provide the solution to the inverse spectral problem. More generally, we would like to do the analysis of (1) from the point of view of [12]. One relevant observation here is that, if (α, β) is a ψ -gap and μ is in $[\alpha, \beta]$, then there is a ξ such that μ is in the spectrum of the multipoint problem for $a_\xi(x)$ and $\rho_\xi(x)$. This is because, if $\lambda \in [\alpha, \beta]$, then (1) always has a Floquet solution, say $f_j(x; \lambda)$, that vanishes for some $x = x_0$. Taking $\xi = x_0$ does the job.

We believe that the same is true for the gaps of the spectrum; namely, that for any λ in a gap, there is a ξ so that λ is in the spectrum of the multipoint problem for some $a_\xi(x)$ and $\rho_\xi(x)$.

The last theorem of this article is an application of Theorems 6 and 7. It completes Theorem 5, so that the two theorems together form the analogue of Theorem B for the pseudospectrum.

Theorem 8. Let $\nu^* = \nu'_{2n-1} = \nu'_{2n}$ or $\nu^* = \nu_{2n-1} = \nu_{2n}$; namely, ν^* is a collapsed (i.e., closed) ψ -gap of (1) (equivalently, ν^* is a double zero of $E(\lambda)$ —see Theorem

D). Then the Floquet matrix $T(\nu^*)$ is similar to the diagonal matrix

$$\begin{pmatrix} r_1 & 0 & 0 & 0 \\ 0 & r_1 & 0 & 0 \\ 0 & 0 & r_1^{-1} & 0 \\ 0 & 0 & 0 & r_1^{-1} \end{pmatrix}.$$

In other words, for the equation

$$[a(x)u''(x)]'' = \nu^* \rho(x)u(x),$$

we have coexistence of four Floquet solutions, two with multiplier r_1 and two with multiplier r_1^{-1} .

Proof. Consider the shifts $a_\xi(x)$ and $\rho_\xi(x)$ of $a(x)$ and $\rho(x)$ of (65). As we observed in Remark 9, the equation

$$(66) \quad [a_\xi(x)u''(x)]'' = \lambda \rho_\xi(x)u(x)$$

has the same spectrum and pseudospectrum as (1), for all $\xi \in \mathbb{R}$. In particular, ν^* is a closed ψ -gap of (66), for all $\xi \in \mathbb{R}$. The Floquet multipliers $r_j = r_j(\nu^*)$, $j = 1, 2, 3, 4$, are the same for all $\xi \in \mathbb{R}$ and we have

$$r_1 = r_2 = r_3^{-1} = r_4^{-1}.$$

Next we consider the multipoint eigenvalue problem

$$(67) \quad [a_\xi(x)u''(x)]'' = \lambda \rho_\xi(x)u(x), \quad u(0) = u(b) = u(2b) = u(3b) = 0.$$

From the above discussion and Theorem 7 it follows that ν^* is a double eigenvalue of (67), for all $\xi \in \mathbb{R}$. Thus, by Theorem 6, for any $\xi \in \mathbb{R}$, there are two Floquet solutions, $f_{1,\xi}(x; \nu^*)$ and $f_{4,\xi}(x; \nu^*)$ of (66), with corresponding multipliers r_1 and $r_4 = r_1^{-1}$ such that

$$(68) \quad f_{1,\xi}(0; \nu^*) = f_{4,\xi}(0; \nu^*) = 0.$$

Each Floquet solution of (66) is a ξ -shift of a Floquet solution $f_j(x; \lambda)$ of (1):

$$f_{1,\xi}(x; \nu^*) = f_1(x + \xi; \nu^*) \quad \text{and} \quad f_{4,\xi}(x; \nu^*) = f_4(x + \xi; \nu^*).$$

Hence, if we did not have coexistence, then (68) would imply

$$f_1(\xi; \nu^*) = f_4(\xi; \nu^*) = 0, \quad \text{for all } \xi \in \mathbb{R},$$

which is impossible since the Floquet solutions considered are nontrivial. Thus we have coexistence of two (linearly independent) Floquet solutions with multiplier r_1 , and two Floquet solutions with multiplier r_1^{-1} . \square

Remark 10. The above theorem, together with Theorems 6 and 7, imply, in particular, that $m_g(\mu) = m_a(\mu)$, for any eigenvalue μ of (50) (see Remark 6).

In relation to Theorem 8 we notice that, if for some ν^* the Floquet matrix $T(\nu^*)$ is similar to the diagonal matrix $\text{diag}(r_1, r_1, r_1^{-1}, r_1^{-1})$, then $T(\nu^*)$ has a very special structure. Its sixteen entries must satisfy various relations.

Finally, we want to mention a conjecture and three open questions.

Conjecture. If all nonzero zeros of $E(\lambda)$ are double (see Theorem D), then $\rho(x)a(x) \equiv 1$. Equivalently: if all the ψ -gaps are closed, then the beam operator is a perfect square of a Hill-type operator.

Open Question 1. An interesting question is: What can be said about $a(x)$ and $\rho(x)$ if we know that the spectrum $S(a, \rho)$ of (1) has no gaps; namely, if $S(a, \rho) = [0, \infty)$?

Open Question 2. In the Hill case there is a simple correspondence between periodic inverse spectral data and inverse spectral data of the separated boundary value problem on the interval $(0, b)$, where b is the period of the potential (see [14]). Find the analogous correspondence for the beam operator. This will be a major step in the solution of the inverse periodic spectral problem for the beam.

Open Question 3. Extend the results presented in this paper to n -th order operators.

APPENDIX

We present here a proposition which was used in the proofs of some of our theorems.

Proposition. For $a_n, \rho_n \in C^\infty[0, b]$, $a_n(x), \rho_n(x) \geq m > 0$, $n = 1, 2, 3, \dots$, consider the initial value problems

(69) $[a_n(x)u_n''(x)]'' = \lambda\rho_n(x)u_n(x), \quad 0 < x < b,$

(70) $u_n(0; \lambda) = \alpha, \quad u_n'(0; \lambda) = \beta, \quad u_n''(0; \lambda) = \gamma, \quad u_n'''(0; \lambda) = \delta,$

where primes denote derivatives with respect to x and $\lambda \in \mathbb{C}$ is a parameter. If

$$a_n(x) \xrightarrow{C^2} a(x) \quad \text{and} \quad \rho_n(x) \xrightarrow{C} \rho(x), \quad x \in [0, b],$$

then

$$u_n(b; \lambda) \rightarrow u(b; \lambda),$$

uniformly on compact subsets of \mathbb{C} , where

$$[a(x)u''(x)]'' = \lambda\rho(x)u(x),$$

$$u(0; \lambda) = \alpha, \quad u'(0; \lambda) = \beta, \quad u''(0; \lambda) = \gamma, \quad u'''(0; \lambda) = \delta.$$

Proof. We write (69) as a first-order system

(71) $\frac{dy_n}{dx} = A_n(x; \lambda)y_n, \quad 0 < x < b,$

where

$$y_n(x; \lambda) = \begin{pmatrix} u_n(x; \lambda) \\ u_n'(x; \lambda) \\ a_n(x)u_n''(x; \lambda) \\ [a_n(x)u_n''(x; \lambda)]' \end{pmatrix}, \quad A_n(x; \lambda) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & a_n(x)^{-1} & 0 \\ 0 & 0 & 0 & 1 \\ \lambda\rho_n(x) & 0 & 0 & 0 \end{pmatrix}.$$

The initial conditions (70) become

$$y_n(0; \lambda) = \begin{pmatrix} \alpha \\ \beta \\ a_n(0)\gamma \\ a_n'(0)\gamma + a_n(0)\delta \end{pmatrix}.$$

Consider also the problem

(72) $\frac{dy}{dx} = A(x; \lambda)y, \quad 0 < x < b,$

(73) $y(0; \lambda) = y_n(0; \lambda),$

where

$$A(x; \lambda) = \lim_n A_n(x; \lambda) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & a(x)^{-1} & 0 \\ 0 & 0 & 0 & 1 \\ \lambda \rho(x) & 0 & 0 & 0 \end{pmatrix}.$$

Let us assume that $|\lambda| \leq M$, where M is a fixed positive number. If we set

$$w_n(x; \lambda) = y_n(x; \lambda) - y(x; \lambda),$$

then by (71), (72), and (73),

$$\begin{aligned} \frac{dw_n}{dx} &= A_n y_n - A y = A_n w_n + (A_n - A) y, & 0 < x < b, \\ w_n(0; \lambda) &= 0. \end{aligned}$$

It follows that

$$w_n(x; \lambda) = \int_0^x A_n(\xi; \lambda) w_n(\xi; \lambda) d\xi + \int_0^x [A_n(\xi; \lambda) - A(\xi; \lambda)] y(\xi; \lambda) d\xi.$$

Thus

$$(74) \quad |w_n(x; \lambda)| \leq \varepsilon_n + \int_0^x \|A_n(\xi; \lambda)\| |w_n(\xi; \lambda)| d\xi,$$

where

$$\varepsilon_n = \int_0^b \|A_n(\xi; \lambda) - A(\xi; \lambda)\| |y(\xi; \lambda)| d\xi.$$

Notice that

$$(75) \quad \lim_n \varepsilon_n = 0.$$

Applying the Gronwall inequality (see [9], Ch. 1, Prob. 1) to (74) we get

$$|w_n(b; \lambda)| \leq \varepsilon_n + \varepsilon_n \int_0^b \|A_n(\xi; \lambda)\| e^{\int_\xi^b \|A_n(x; \lambda)\| dx} d\xi$$

and $\|A_n(x; \lambda)\|$ is bounded for $x \in [0, b]$ and $|\lambda| \leq M$ (uniformly in n). Therefore, by (75) we have that

$$w_n(b; \lambda) = y_n(b; \lambda) - y(b; \lambda) \rightarrow 0$$

uniformly in λ , as long as $|\lambda| \leq M$. Since M is arbitrary, the proof is complete. \square

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professors Peter Kuchment and Sergey Novikov for helpful discussions and suggested references. We also want to thank Professor George Papanicolaou and the (NSF-funded) Mathematical Geophysics Summer School, held at Stanford University, for the great hospitality and the partial support of this work. Finally, the author thanks the anonymous referee for numerous constructive suggestions and comments.

REFERENCES

- [1] J. E. AVRON AND B. SIMON, Analytic Properties of Band Functions, *Annals of Physics*, **110** (1978), 85–101. MR **57**:14992
- [2] A. BADANIN AND E. KOROTYAEV, Quasimomentum of Fourth Order Periodic Operator, preprint, 2001.
- [3] V. BARCILON, Inverse Problem for a Vibrating Beam in the Free-Clamped Configuration, *Philosophical Transactions of the Royal Society of London, Series A*, **304** (1982), 211–251.
- [4] R. BEALS AND R. R. COIFMAN, Scattering and Inverse Scattering for First Order Systems, *Comm. Pure Appl. Math.* **37**, no. 1 (1984), 39–90. MR **85f**:34020
- [5] R. BEALS, P. DEIFT, AND C. TOMEI, Direct and Inverse Scattering on the Line, *Mathematical Surveys and Monographs*, **28**. American Mathematical Society, Providence, RI, 1988, xiv + 209 pp. MR **90a**:58064
- [6] R. CARLSON, Compactness of Floquet Isospectral Sets for the Matrix Hill's Equation, *Proceedings of the American Mathematical Society* **128**, no. 10 (2000), 2933–2941. MR **2000m**:34027
- [7] R. CARLSON, Eigenvalue Estimates and Trace Formulas for the Matrix Hill's Equation, *Journal of Differential Equations* **167**, no. 1 (2000), 211–244. MR **2001e**:34157
- [8] L. F. CAUDILL, P. A. PERRY, AND A. W. SCHUELLER, Isospectral Sets for Fourth-Order Ordinary Differential Operators, *SIAM J. Math. Anal.* **29** (1998), 935–966. MR **99c**:34022
- [9] E. A. CODDINGTON AND N. LEVINSON, "Theory of Ordinary Differential Equations", Robert E. Krieger Publishing Company, Malabar, Florida, 1987. MR **16**:1022b
- [10] W. CRAIG, The Trace Formula for Schrödinger Operators on the Line, *Commun. Math. Phys.* **126** (1989), 379–407. MR **90m**:47063
- [11] B. A. DUBROVIN, I. M. KRICHEVER, AND S. P. NOVIKOV, Integrable Systems. I, *Dynamical Systems, IV*, 177–332, *Encyclopaedia of Mathematical Sciences*, **4**, Springer-Verlag, Berlin, 2001. MR **87k**:58112
- [12] B. A. DUBROVIN, V. MATVEEV, AND S. P. NOVIKOV, Nonlinear Equations of Korteweg-deVries Type, Finite Zone Linear Operators, and Abelian Varieties, *Uspekhi. Mat. Nauk*, **31** (1976), 55–136; *Russian Math. Surveys*, **31** (1976), 59–146. MR **55**:899
- [13] N. DUNFORD AND J. T. SCHWARTZ, "Linear Operators. Part II: Spectral Theory; Self Adjoint Operators in Hilbert Space", Wiley Classics Library Edition, New York, 1988. MR **90g**:47001b
- [14] A. FINKEL, E. ISAACSON, AND E. TRUBOWITZ, An Explicit Solution of the Inverse Periodic Problem Hill's Equation, *SIAM J. Math. Anal.* **18**, No. 1 (Jan. 1987), 46–53. MR **88d**:34037
- [15] F. GESZTESY, H. HOLDEN, B. SIMON, AND Z. ZHAO, Trace Formulae and Inverse Spectral Theory for Schrödinger Operators, *Bull. Amer. Math. Soc. (New Series)* **29** (1993), 250–255. MR **94c**:34127
- [16] F. GESZTESY AND R. WEIKARD, Floquet Theory Revisited, in "Differential Equations and Mathematical Physics", Proceedings of the International Conference, Univ. of Alabama at Birmingham, March 13–17, 1994, International Press, 1995. MR **2000i**:34163
- [17] G. M. L. GLADWELL, "Inverse Problems in Vibration," Martinus Nijhoff Publishers, Boston, 1986. MR **88b**:73002
- [18] R. JOHNSON, m -Functions and Floquet Exponents for Linear Differential Systems, *Annali di Matematica pura ed applicata (IV)*, **CXLVII** (1987), 211–248. MR **88m**:34021
- [19] R. JOHNSON AND J. MOSER, The Rotation Number for Almost Periodic Potentials, *Comm. Math. Phys.* **84** (1982), 403–438; erratum: *Comm. Math. Phys.* **90** (1983), 317–318. MR **83h**:34018
- [20] W. KOHN, Analytic Properties of Bloch Waves and Wannier Functions, *Annals of Physical Review*, **115**, no. 4 (August 1959), 809–821. MR **21**:7000
- [21] P. KUCHMENT, "Floquet Theory for Partial Differential Equations", Birkhäuser-Verlag, Basel, 1993. MR **94h**:35002
- [22] W. MAGNUS AND S. WINKLER, "Hill's Equation", Dover Publications, Inc., New York, 1979. MR **80k**:34001
- [23] M. M. MALAMUD, Necessary Conditions for the Existence of a Transformation Operator for Higher Order Equations, *Funktsional. Anal. i Ego Philozhen.* **16** (1982), 74–75. *Functional Anal. Appl.* **16** (1982), 219–221. MR **84i**:34011

- [24] H. P. MCKEAN AND E. TRUBOWITZ, Hill's Operator and Hyperelliptic Function Theory in the Presence of Infinitely Many Branch Points, *Comm. Pure Appl. Math.* **29**, no. 2 (1976), 143–226. MR **55**:761
- [25] H. P. MCKEAN AND P. VAN MOERBEKE, The Spectrum of Hill's Equation, *Invent. Math.* **30**, no. 3 (1975), 217–274. MR **53**:936
- [26] J. R. McLAUGHLIN, On Constructing Solutions to an Inverse Euler-Bernoulli Problem, in "Inverse Problems of Acoustic and Elastic Waves", pp. 341–347, F. Santosa, et al. (editors), Philadelphia: SIAM, 1984. MR **86e**:00016
- [27] J. R. McLAUGHLIN, Analytical Methods for Recovering Coefficients in Differential Equations from Spectral Data, *SIAM Review* **28** (1986), 53–72. MR **87d**:34034
- [28] R. E. MILLER, The Eigenvalue Problem for a Class of Long, Thin Elastic Structures with Periodic Geometry, *Quarterly of Applied Mathematics*, **LII**, No. 2 (June 1994), 261–282. MR **95c**:73008
- [29] M. A. NAIMARK, "Linear Differential Operators", Parts I & II, Frederick Ungar Publishing Co., New York, 1967 & 1968. MR **35**:6885, MR **41**:7485
- [30] S. P. NOVIKOV, private communication (April 2001).
- [31] V. G. PAPANICOLAOU, The Spectral Theory of the Vibrating Periodic Beam, *Comm. Math. Phys.* **170** (1995), 359–373. MR **96d**:34108
- [32] V. G. PAPANICOLAOU AND D. KRAVVARITIS, An Inverse Spectral Problem for the Euler-Bernoulli Equation for the Vibrating Beam, *Inverse Problems* **13** (1997), 1083–1092. MR **98f**:34016
- [33] V. G. PAPANICOLAOU AND D. KRAVVARITIS, The Floquet Theory of the Periodic Euler-Bernoulli Equation, *Journal of Differential Equations* **150** (1998), 24–41. MR **2000a**:34167
- [34] M. REED AND B. SIMON, "Methods of Modern Mathematical Physics, VI: Analysis of Operators", Academic Press, New York, 1978. MR **58**:12429c
- [35] L. A. SACHNOVICH, Inverse Problems for Differential Equations of Order $n > 2$ with Analytic Coefficients, *Matematicheskii Sbornik* **46** (1958), 61–76. MR **20**:5912
- [36] S. TIMOSHENKO AND D. H. YOUNG, "Elements of Strength of Materials", 5th Edition, D. Van Nostrand Company, Inc., Princeton, NJ, 1968.
- [37] E. C. TITCHMARSH, "The Theory of Functions", Second Edition, Oxford University Press, 1939.
- [38] E. TRUBOWITZ, The Inverse Problem for Periodic Potentials, *Comm. Pure Appl. Math.* **30**, no. 3 (1977), 321–337. MR **55**:3408
- [39] S. VENAKIDES, private communication.

DEPARTMENT OF MATHEMATICS AND STATISTICS, WICHITA STATE UNIVERSITY, WICHITA, KANSAS 67260-0033

Current address: Department of Mathematics, National Technical University of Athens, Zografou Campus, 157 80, Athens, Greece

E-mail address: papanico@math.ntua.gr

SINGULARITIES OF THE HYPERGEOMETRIC SYSTEM ASSOCIATED WITH A MONOMIAL CURVE

FRANCISCO JESÚS CASTRO-JIMÉNEZ AND NOBUKI TAKAYAMA

ABSTRACT. We compute, using \mathcal{D} -module restrictions, the slopes of the irregular hypergeometric system associated with a monomial curve. We also study rational solutions and reducibility of such systems.

1. INTRODUCTION

Let $A_n = \mathbf{C}\langle x_1, \dots, x_n, \partial_1, \dots, \partial_n \rangle$ be the Weyl algebra of order n over the complex numbers \mathbf{C} and let $\mathbf{C}[\partial] = \mathbf{C}[\partial_1, \dots, \partial_n]$ be the subring of A_n of linear differential operators with constant coefficients.

Let $A = (a_{ij})$ be an integer $(d \times n)$ -matrix of rank d . We denote by $I_A \subset \mathbf{C}[\partial]$ the *toric ideal* associated to A : i.e., I_A is the ideal generated by the set

$$\{\partial^u - \partial^v \mid u, v \in \mathbf{N}^n, Au^T = Av^T\}$$

where $()^T$ means “transpose”.

We denote by θ the vector $(\theta_1, \dots, \theta_n)^T$ with $\theta_i = x_i \partial_i$. For a given $\beta = (\beta_1, \dots, \beta_d)^T \in \mathbf{C}^d$ we consider the column vector (in A_n^d) $A\theta - \beta$ and we denote by $\langle A\theta - \beta \rangle$ the left ideal of A_n generated by the entries of $A\theta - \beta$.

Following Gel’fand, Zelevinskii and Kapranov [6], we denote by $H_A(\beta)$ the left ideal of A_n generated by $I_A \cup \langle A\theta - \beta \rangle$. It is called the GKZ-hypergeometric system associated to the pair (A, β) . The quotient $\mathcal{H}_A(\beta) = A_n / H_A(\beta)$ is a holonomic A_n -module (see e.g. [17]).

If the toric ideal I_A is homogeneous, i.e., if the \mathbf{Q} -row span of A contains $(1, \dots, 1)$, it is known ([8]; see also [17]) that $\mathcal{H}_A(\beta)$ is regular holonomic and the book [17] is devoted to an algorithmic study of such systems. Especially, the book gives an algorithmic method to construct series solutions around singular points of the system.

In this article, we start a study of singularities of GKZ-hypergeometric systems for non-homogeneous toric ideals I_A by treating the “first” case when $d = 1$, $A = (a_1, a_2, \dots, a_n) \in \mathbf{Z}^n$, $a_1 = 1$. We evaluate the geometric slopes of $\mathcal{H}_A(\beta)$ by successive restrictions of the number of variables. The slopes characterize Gevrey class solutions around a singular locus. Let us remark on an analytic meaning of slopes. Let $X = \mathbf{C}^n$ and $Y = \{x \mid x_n = 0\} \subset X$. We denote by $\mathcal{O}_{\widehat{X|Y}}$ the

Received by the editors November 15, 2002.

2000 *Mathematics Subject Classification*. Primary 32C38, 13N10; Secondary 13P10, 14F10, 14M25.

Key words and phrases. Algebraic geometry, \mathcal{D} -modules, toric varieties, hypergeometric systems.

The first author was partially supported by BFM-2001-3164, FQM-218 and FQM-813.

formal completion of \mathcal{O}_X along Y (i.e., series that are formal in x_n and convergent in x_1, \dots, x_{n-1}). To each real number $s \in [1, +\infty)$ we denote by $\mathcal{O}_{X|Y}(s)$ the subsheaf of $\widehat{\mathcal{O}_{X|Y}}$ of Gevrey functions of order s (along Y). The sheaf $\mathcal{O}_{X|Y}(1)$ is the restriction $\mathcal{O}_{X|Y}$ and, by definition, we write $\mathcal{O}_{X|Y}(+\infty) = \widehat{\mathcal{O}_{X|Y}}$. For any holonomic \mathcal{D}_X -module M , Z. Mebkhout [14] associates the sheaf $\text{Irr}_Y(s)(M)$ as the solution sheaf $\mathbf{R}Hom_{\mathcal{D}}(M, \mathcal{O}_{X|Y}(s)/\mathcal{O}_{X|Y})$. One fundamental result in the irregularity of \mathcal{D} -modules is the fact that $\text{Irr}_Y(s)(M)$ is a perverse sheaf, for any s (see [14]). These sheaves define a filtration of the irregularity of M along Y , i.e., $\text{Irr}_Y(M) := \text{Irr}_Y(+\infty)(M)$. The main result of [11] is that $1/(1-s)$ is a slope of M with respect to Y if and only if s is a gap of the graduation defined by the filtration on the irregularity. In other words, $1/(1-s)$ is a slope if and only if $\text{Irr}_Y(s)(M)/\text{Irr}_Y(<s)(M) \neq 0$.

Our evaluation of the slopes is done as follows: (1) We translate Laurent and Mebkhout's theorem [12] on restrictions and slopes of \mathcal{D} -modules into an algorithm to evaluate the slopes by utilizing the results of [2] and [15]. (2) Apply our general algorithm to the hypergeometric system associated to $A = (1, a_2, \dots, a_n)$. This system has many nice properties and our algorithm outputs the slopes without computation on computers.

In the last section we study rational solutions and reducibility of our systems.

2. MICRO-CHARACTERISTIC VARIETIES

In this section, following Laurent [10], we describe micro-characteristic varieties for a given \mathcal{D} -module. We will state a result of Laurent and Mebkhout [12, Corollaire 2.2.9] (see also [14, p. 125] and [11, p. 42]), allowing to reduce our general problem of evaluating the slopes to fewer variables.

In this section $X = \mathbf{C}^n$ and $\mathcal{D}_X = \mathcal{D}$ is the sheaf of linear differential operators with holomorphic function coefficients. Let M be a coherent \mathcal{D} -module. Recall that the characteristic variety of M (denoted by $\text{Ch}(M)$) is an analytic subvariety of the cotangent bundle T^*X .

Suppose that $Y \subset X$ is a smooth hypersurface. We say that Y is *non-characteristic* for M if $T_Y^*X \cap \text{Ch}(M) \subset T_X^*X$. Here T_Y^*X is the conormal bundle to Y in X and T_X^*X is the zero section of T^*X .

Now, following Laurent [9], [10], we shall define the notion of *non-micro-characteristic variety* for M . To simplify the presentation we will assume that (x_1, x_2, \dots, x_n) are local coordinates in X and that Y is defined by $x_n = 0$. We denote by $(x_1, \dots, x_n, \xi_1, \dots, \xi_n)$ local coordinates in T^*X . Sometimes it will be useful to write $x_1 = y_1, \dots, x_{n-1} = y_{n-1}$, $x_n = t$, $\xi_1 = \eta_1, \dots, \xi_{n-1} = \eta_{n-1}$ and $\xi_n = \tau$.

Let us denote by Λ the conormal bundle of Y in T^*X (i.e., $\Lambda = T_Y^*X$). So, in local coordinates, $\Lambda = \{(y, t, \eta, \tau) \in T^*X \mid t = \eta = 0\}$. Here, $\eta = (\eta_1, \dots, \eta_{n-1})$ and $y = (y_1, \dots, y_{n-1})$. We denote by (y, τ, y^*, τ^*) local coordinates on the cotangent bundle $T^*\Lambda$.

We denote by $V_{\bullet}(\mathcal{D})$ (or simply by V) the Malgrange-Kashiwara filtration associated to Y on \mathcal{D} and by $F_{\bullet}(\mathcal{D})$ (or simply by F) the order filtration on \mathcal{D} . For a given rational number $p/q \geq 0$ we denote by $L_{p/q}$ the filtration on \mathcal{D} defined by $pF + qV$. The $L_{p/q}$ -order of a monomial $y^{\alpha} t^l \partial_y^{\beta} \partial_t^k$ is equal to $p(|\beta| + k) + q(k - l)$, where $|\beta| = \beta_1 + \dots + \beta_{n-1}$. We will simply write $L = L_{p/q}$ if no confusion arises.

For $p > 0$ the associated graded ring $\text{gr}^L(\mathcal{D})$ is canonically isomorphic to $\pi_* \mathcal{O}_{[T^*\Lambda]}$ [10, p. 407], where $\pi : T^*\Lambda \rightarrow \Lambda$ is the canonical projection and $\mathcal{O}_{[T^*\Lambda]}$ denotes

holomorphic functions on $T^*\Lambda$ that are polynomials on the fibers of π . In local coordinates, $\text{gr}^L(\mathcal{D}_0)$ is expressed as $\mathbf{C}\{y\}[\tau, y^*, \tau^*]$ where \mathcal{D}_0 is the stalk of \mathcal{D} at the origin.

Given a differential operator

$$P = \sum_{\alpha l \beta k} p_{\alpha l \beta k} y^\alpha t^l \partial_y^\beta \partial_t^k,$$

the L -order of P is the maximum value of $p(|\beta| + k) + q(k - l)$ over the monomials of P . For $p > 0$ we define the L -principal symbol of P by

$$\sigma^L(P) = \sum p_{\alpha l \beta k} y^\alpha (\tau^*)^l (y^*)^\beta (-\tau)^k$$

where the sum is taken over monomials with maximal L -order. The L -principal symbol of P is an element of $\text{gr}^L(\mathcal{D})$ and then is a function on $T^*\Lambda$. In the classical case, i.e., for $L = F$, $\text{gr}^F(\mathcal{D})$ is identified with $\mathbf{C}\{x\}[\xi_1, \dots, \xi_n] = \mathbf{C}\{y, t\}[\eta, \tau]$ and the F -principal symbol of P is simply denoted by $\sigma^F(P) = \sum p_{\alpha l \beta k} y^\alpha t^l \eta^\beta \tau^k$ where the sum is taken for $|\beta| + k$ maximum.

To each left ideal $I \subset \mathcal{D}$ we denote by $\sigma^L(I)$ the ideal of $\text{gr}^L(\mathcal{D})$ generated by the set of $\sigma^L(P)$ for $P \in I$.

For each L -filtration on \mathcal{D} we associate a “good” L -filtration on M , by means of a finite presentation. The associated $\text{gr}^L(\mathcal{D})$ -module $\text{gr}^L(M)$ is coherent (see [10, 3.2.2]). The radical of the annihilating ideal $\text{Ann}_{\text{gr}^L(\mathcal{D}_X)}(\text{gr}^L(M))$, which is independent of the “good” filtration on M , defines an analytic subvariety of $T^*\Lambda$. This variety is called the L -characteristic variety of M and it is denoted by $\text{Ch}^L(M)$.

Suppose now that $Z \subset X$ is a smooth hypersurface transverse to Y . Suppose for simplicity that Z is defined in local coordinates by $y_1 = 0$. The conormal space $\Lambda' := T_{Y \cap Z}^*Z$ is a smooth subvariety of $\Lambda = T_Y^*X$ defined in local coordinates by $y_1 = 0$. So $T_{\Lambda'}^*\Lambda$ is the subvariety of $T^*\Lambda$ defined in local coordinates by $y_1 = y'^* = \tau^* = 0$, where $y' = (y_2, \dots, y_{n-1})$.

Definition 2.1 ([9], [12]). We say that Z is *non-micro-characteristic of type L* for M if $T_{\Lambda'}^*\Lambda \cap \text{Ch}^L(M)$ is contained in $T_{\Lambda'}^*\Lambda$. Sometimes we will say that, if this condition holds, Z is *non- L -micro-characteristic* for M .

The sheaf of rings $\text{gr}^L(\mathcal{D})$ is endowed with two graduations: The first is induced by the F -filtration and the second is induced by the V -filtration. Recall Laurent’s definition of slope of a coherent \mathcal{D} -module M .

Definition 2.2 ([10]). The rational number $-p/q$ is said to be a *slope* of M with respect to Y at the origin if and only if the radical of the ideal $\text{Ann}_{\text{gr}^L(\mathcal{D}_X)}(\text{gr}^L(M))$ is not bihomogeneous for F - nor V -graduations.

An important consequence of the work [11] (see Théorème 2.4.2) is what follows: A holonomic \mathcal{D}_X -module M is regular with respect to Y at the origin if and only if M has no slope with respect to Y at the origin. We will use this fact freely in the text.

Remark. In [12], ∞ and 0 (F and V) were included in the set of the slopes. We do not include them in the set of the slopes in this paper.

Finally, the following result by Laurent and Mebkhout allows induction on the number of variables to calculate slopes.

Theorem 2.3 ([12, Corollaire 2.2.9]). *Let M be a holonomic \mathcal{D}_X -module. Let Z and Y be transverse smooth hypersurfaces on X such that Z is non- L -micro-characteristic (for all $p > 0, q > 0$) for M . Then the slopes of M with respect to Y equal the slopes of M' with respect to $Y' = Z \cap Y$, where M' is the restriction of M to Z .*

This is a deep result in \mathcal{D} -module theory. Its proof uses the algebraic-analytic comparison theorem ([11, Theorem 2.4.2]) and a Cauchy-Kowalewska theorem for Gevrey functions with respect to Z ([12, Corollaire 2.2.4]; see also [14, Théorème 6.3.4]).

3. COMPUTING SLOPES BY REDUCING THE NUMBER OF VARIABLES

We have introduced the notion of the slopes and the invariance of them under restrictions satisfying a condition on L -characteristic varieties. In the sequel, we assume that our ideal is that of the Weyl algebra A_n . Constructions in sheaves such as restrictions and L -characteristic varieties in the previous section can be done via constructions in the Weyl algebra as we usually see in the computational \mathcal{D} -module theory.

We are interested in computation of the slopes. The slopes of A_n/I (at the origin) along $x_n = 0$ can be computed by the ACG algorithm introduced in [2]. In this section, translating Laurent and Mebkhout's result into computer algebra algorithms, we will give a preprocessing method for the ACG algorithm to accelerate the original. The preprocessing is useful for a class of inputs including GKZ hypergeometric ideals as we will see in Section 4. Let us first recall the ACG algorithm.

A weight vector is an element $W = (u_1, \dots, u_n, v_1, \dots, v_n) \in \mathbf{R}^{2n}$ such that $u_i + v_i \geq 0$ for all i . This weight vector $W = (u, v)$ induces a natural filtration on A_n , and it is called the W -filtration. The associated graded ring is denoted by $\text{gr}^W(A_n)$, and for each left ideal $I \subset A_n$ the associated graded ideal is denoted by $\text{in}_W(I)$ or $\text{in}_{(u,v)}(I)$. Here, the *initial ideal* $\text{in}_{(u,v)}(I)$ is the ideal generated by $\text{in}_{(u,v)}(f)$, $f \in I$ in $\text{gr}^W(A_n)$. When $u_i + v_i > 0$, it is an ideal in the polynomial ring of $2n$ variables: $\text{gr}^W(A_n) = \mathbf{C}[x_1, \dots, x_n, \xi_1, \dots, \xi_n]$. The initial ideal of I with respect to the weight (u, v) is generated by the (u, v) -initial terms of a Gröbner basis of I by an order that refines the partial order defined by (u, v) . See, e.g., [17, Theorem 1.1.6].

Consider the filtration $L = pF + qV$, $p > 0, q > 0$ introduced in the previous section. The ideal $\sigma^L(I)$, which gives the L -characteristic variety, can be expressed in terms of the initial ideal as follows:

$$\sigma^L(I) = \text{gr}^L(\mathcal{D}) \cdot \text{in}_\ell(I) |_{x_1 \mapsto y_1, \dots, x_{n-1} \mapsto y_{n-1}, x_n \mapsto \tau^*, \xi_1 \mapsto y_1^*, \dots, \xi_{n-1} \mapsto y_{n-1}^*, \xi_n \mapsto -\tau},$$

$$\ell = p(\overbrace{0, \dots, 0}^n, \overbrace{1, \dots, 1}^n) + q(\overbrace{0, \dots, 0}^n, \overbrace{-1, 0, \dots, 0}^n, 1).$$

For two weight vectors W and W' and a term order $<$, we denote by $<_{W,W'}$ the order

$$\begin{aligned} x^\alpha \partial^\beta &<_{W,W'} x^a \partial^b \\ \Leftrightarrow W \cdot (\alpha, \beta) &< W \cdot (a, b) \\ \text{or } W \cdot (\alpha, \beta) &= W \cdot (a, b) \text{ and } W' \cdot (\alpha, \beta) < W' \cdot (a, b) \\ \text{or } W^* \cdot (\alpha, \beta) &= W^* \cdot (a, b) \text{ for both } W^* = W, W' \text{ and } x^\alpha \partial^\beta < x^a \partial^b. \end{aligned}$$

To each differential operator $P = \sum p_{\alpha\beta} x^\alpha \partial^\beta \in A_n$ we associate the Newton polygon $N(P)$ of P (with respect to $x_n = 0$) defined as the convex hull of the subset of \mathbf{Z}^2 ,

$$\bigcup_{p_{\alpha\beta} \neq 0} (|\beta|, \beta_n - \alpha_n) + (-\mathbf{N})^2.$$

Let I be a left ideal in A_n . As we said in Definition 2.2, the notion of slope of a differential system was introduced by Y. Laurent [10]. Let us give here a slightly different but equivalent definition: the number r , $-\infty < r < 0$, is a *geometric* slope of I (or of A_n/I) with respect to $x_n = 0$ if and only if $\sqrt{\sigma^{(-r)F+V}(I)}$ is not bihomogeneous with respect to the weight vectors $F = (0, \dots, 0, 1, \dots, 1)$ and $V = (0, \dots, 0, -1, 0, \dots, 0, 1)$. Following [2], we say that the number r , $-\infty < r < 0$ is an *algebraic* slope of I (or of A_n/I) if and only if $\sigma^{(-r)F+V}(I)$ is not bihomogeneous with respect to the weight vectors F and V . The geometric slope is simply called the slope in this paper if confusion does not arise. Note that we may consider $\text{in}_L(I)$ instead of $\sigma^L(I)$ so long as we are concerned about homogeneity. For algebraic or geometric slope r , the weight vector $L = (-r)F + V$ lies on a face of the Gröbner fan of I ([3], [17]), which yields the following algorithm.

Algorithm 3.1 ([2], ACG algorithm).

Input: $G = \{P_1, \dots, P_m\}$ (generators of an ideal I).

Output: All algebraic and geometric slopes of A_n/I with respect to $x_n = 0$ at the origin.

```

geometric_slope =  $\emptyset$ ; algebraic_slope =  $\emptyset$ ;
F = (0, ..., 0, 1, ..., 1); V = (0, ..., 0, -1, 0, ..., 0, 1);
p = 1; q = 0; slope =  $-\infty$ ; previous_slope = slope;
while (slope  $\neq$  0) {
    L = pF + qV;
    G = a Gröbner basis of  $I$  with respect to the order  $<_{L,V}$ ;
    slope = the minimum of 0 and
        {the slopes  $r$  of the Newton polygon  $N(P) \mid P \in G, r > \text{previous\_slope}$ }
    if slope = 0, then return (algebraic_slope and geometric_slope).
    if  $\sigma^L(G)$  is not homogeneous for  $F$  nor  $V$  then {
        algebraic_slope = algebraic_slope  $\cup$  {slope}
    }
    if  $\sqrt{\sigma^L(G)}$  is not homogeneous for  $F$  nor  $V$  then {
        geometric_slope = geometric_slope  $\cup$  {slope}
    }
    p = numerator(|slope|); q = denominator(|slope|);
    previous_slope = slope;
}
return(algebraic_slope and geometric_slope)

```

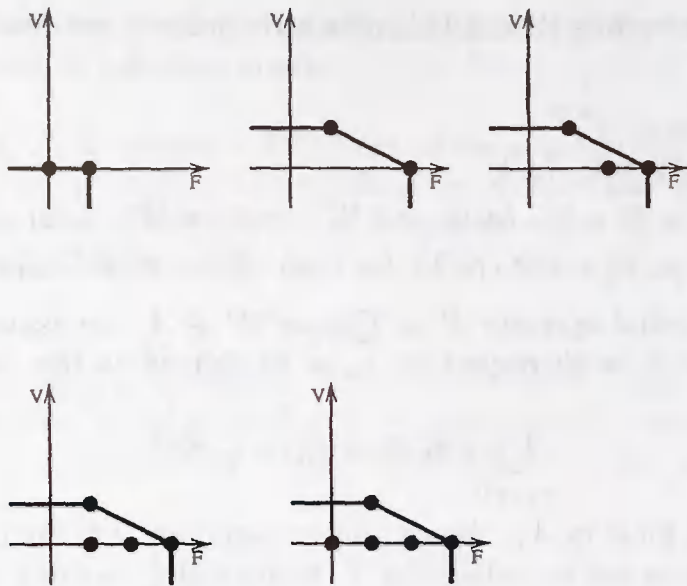


FIGURE 1

Here, we use the convention $F := \infty F + V$. Note that the ideal $J = \langle f_1, \dots, f_m \rangle \subset \mathbb{C}[x_1, \dots, x_n, \xi_1, \dots, \xi_n]$ is homogeneous for the weight $(u, v) \in \mathbb{Z}^{2n}$ if and only if all (u, v) -homogeneous subsums of f_i belong to the ideal J .

For ordinary differential equations, the ACG algorithm is nothing but the well-known Newton polygon method; see two examples below.

Example 3.2. Put $G = \{x_1^{p+1}\partial_1 + p\}$, $n = 1$, $p \in \mathbb{N}$. Then, the ACG algorithm returns $\{-p\}$. ($\exp(x_1^{-p})$ is a classical solution of G .)

Example 3.3. Put $G = \{2x_1(x_1\partial_1)^2 + x_1^2\partial_1 + 1\}$, $n = 1$. Then, the ACG algorithm returns $\{-1/2\}$. ($\exp(x_1^{-1/2})$ is a classical solution of G .)

The next example of two variables is generated by the computer algebra system `kan/k0` [18].

Example 3.4. We consider the GKZ hypergeometric ideal I associated to the matrix $A = (1, 3)$ and $\beta = -3$. We will compute the slopes of A_2/I at the origin along $x_2 = 0$ by the ACG algorithm. The Gröbner basis of I for the weight

$$\left(\overbrace{0}^{x_1}, \overbrace{0}^{x_2}, \overbrace{1}^{\partial_1}, \overbrace{1}^{\partial_2} \right) \text{ is}$$

$$\begin{aligned} &[\ x_1 \cdot Dx_1 + 3 \cdot x_2 \cdot Dx_2 + 3, \ -Dx_1^3 + Dx_2, \ -3 \cdot x_2 \cdot Dx_1^2 \cdot Dx_2 - 5 \cdot Dx_1^2 - x_1 \cdot Dx_2, \\ &\quad -9 \cdot x_2^2 \cdot Dx_1 \cdot Dx_2^2 - 36 \cdot x_2 \cdot Dx_1 \cdot Dx_2 + x_1^2 \cdot Dx_2 - 20 \cdot Dx_1, \\ &\quad -27 \cdot x_2^3 \cdot Dx_2^3 - 189 \cdot x_2^2 \cdot Dx_2^2 - x_1^3 \cdot Dx_2 - 276 \cdot x_2 \cdot Dx_2 - 60 \] \end{aligned}$$

Here, Dx_i and x_i stand for ∂_i and x_i , respectively. The Newton polygons $N(P)$'s are given in Figure 1.

From the Newton polygons, $-1/2$ is the candidate of the first slope. Next, we compute the Gröbner basis for the weight $(0, -2, 1, 3) = (0, 0, 1, 1) + 2(0, -1, 0, 1)$. The Gröbner basis is

$$[\ x_1 \cdot Dx_1 + 3 \cdot x_2 \cdot Dx_2 + 3, \ Dx_2 - Dx_1^3 \]$$

The radical is generated by

$$\begin{aligned} &[\ -x_1 \cdot Dx_1 - 3 \cdot x_2 \cdot Dx_2, \ -Dx_1^3 + Dx_2, \ 3 \cdot x_2 \cdot Dx_1^2 \cdot Dx_2 + x_1 \cdot Dx_2, \\ &\quad -9 \cdot x_2^2 \cdot Dx_1 \cdot Dx_2^2 + x_1^2 \cdot Dx_2, \ 27 \cdot x_2^3 \cdot Dx_2^3 + x_1^3 \cdot Dx_2 \] \end{aligned}$$

Here, $x_1 = y_1$, $x_2 = \tau^*$, $Dx_1 = y_1^*$, $Dx_2 = -\tau$. It is not bihomogeneous, and then $-1/2$ is a geometric and algebraic slope.

By looking at the two Newton polygons of $x_1 Dx_1 + 3x_2 Dx_2 + 3$, $Dx_2 - Dx_1^3$, we see that there are no slopes larger than $-1/2$. Then, the ACG algorithm terminates here.

The ACG algorithm requires a repetition of Gröbner basis computations in the Weyl algebra of $2n$ variables to evaluate the slopes of A_n/I . However, if $x_i = 0$ is non-micro-characteristic of type $L = (-r)F + V$ for all $-\infty < r < 0$ and the restriction of A_n/I to $x_i = 0$ is singly generated, we can preprocess the input so that the input ideal for the ACG algorithm lies in the Weyl algebra of $2(n-1)$ variables. The correctness of the following algorithm can be shown by Laurent and Mebkhout's Theorem 2.3.

Algorithm 3.5 (Computing slopes with a preprocessing).

Step 1: Check if $x_i = 0$ is non-micro-characteristic of A_n/I for all types L by calling Algorithm 3.6.

Step 2: Compute the restriction of A_n/I to $x_i = 0$ and check if it is expressed as D'/I' where D' is the Weyl algebra of $2(n-1)$ variables.

Step 3: If we failed either in Step 1 or in Step 2, then apply the ACG algorithm for I .

If we succeeded both in Step 1 and in Step 2, then try to reduce more variables or apply the ACG algorithm for I' .

We can compute restrictions of a given \mathcal{D} -module by using Oaku's algorithm [15]. This algorithm is implemented in computer algebraic systems Macaulay2 and Kan [7], [18], [13]. Therefore, the remaining algorithmic question for the preprocessing is to determine the range of type $L = (-r)F + V$ for which $x_i = 0$ ($i \leq n-1$) is non-micro-characteristic. It follows from the definition of non-micro-characteristic that the question is nothing but to find the segment $(-\infty, r_1)$ such that

$$\mathcal{V}(\sigma^{(-r)F+V}(I), y_1^*, \dots, y_{i-1}^*, y_i, y_{i+1}^*, \dots, y_{n-1}^*, \tau^*) \subseteq \mathcal{V}(y_1^*, \dots, y_{n-1}^*, \tau^*) = T_\Lambda^* \Lambda$$

for $r \in (-\infty, r_1)$. Here, $\mathcal{V}(f_1, \dots, f_m)$ is the affine variety defined by the polynomials f_1, \dots, f_m . The inclusion condition can be algebraically rephrased as

$$\sqrt{\text{in}_{(-r)F+V}(I), \xi_1, \dots, \xi_{i-1}, x_i, \xi_{i+1}, \dots, \xi_n} \ni \xi_i.$$

Since the Gröbner fan is a finite union of Gröbner cones [2], the range can be determined by a similar method with the ACG algorithm.

Algorithm 3.6.

`range_of_nonMC(H, r_0)`

Input: H is a finite set in A_n , r_0 is a negative number or $-\infty$.

Output: r_1 such that $x_1 = 0$ is non-micro-characteristic of type $(-r)F + V$ for $A_n/A_n \cdot \{H\}$, for $r \in [r_0, r_1)$.

`previous_slope = slope = r_0 ;`

$F = (0, \dots, 0, 1, \dots, 1)$; $V = (0, \dots, 0, -1, 0, \dots, 0, 1)$;

while (`slope`! = 0) {

$p = \text{numerator}(|\text{slope}|)$; $q = \text{denominator}(|\text{slope}|)$;

$L = pF + qV$;

$G =$ a Gröbner basis of H with respect to $<_L$;

```

if  $\sqrt{\langle \sigma^L(H), y_1, y_2^*, \dots, y_{n-1}^*, \tau^* \rangle} \ni y_1^*$  and
 $\sqrt{\langle \sigma^{pF+(q+\varepsilon)V}(H), y_1, y_2^*, \dots, y_{n-1}^*, \tau^* \rangle} \ni y_1^*$ 
then {
    previous_slope = slope;
    slope = the minimum of 0 and
        {the slopes  $r$  of the Newton polygon  $N(P) \mid P \in G, r > \text{previous\_slope}$ }
} else {
return(slope);
}
}
return(0);

```

Here, ε is a sufficiently small positive rational number so that $pF + (q + \varepsilon)V$ lies in the interior of a Gröbner cone.

In the case when $r_0 = -\infty$, we use the convention $\text{numerator}(|r_0|) = 1$ and $\text{denominator}(|r_0|) = 0$, and F -non-micro-characteristic means that it is non-characteristic in the classical sense.

Example 3.7. Suppose $n \geq 2$. Then $\text{range_of_nonMC}(\{\partial_1^2 - \partial_2\}, -\infty)$ returns 0.

If the function $\text{range_of_nonMC}(I, -\infty)$ returns 0, then $x_1 = 0$ is non-micro-characteristic of type $pL + qV$ ($p > 0, q \geq 0$) for A_n/I . We note that it is not always a clever strategy to call the function with the full set of generators. In fact, if $x_1 = 0$ is non-micro-characteristic of type L for A_n/J , then it is non-micro-characteristic of type L for A_n/I for any $I \supseteq J$. Therefore, for Step 1, it is sometimes more efficient to call the function range_of_nonMC for a subset of the generators of the input ideal as we will see in the case of GKZ hypergeometric ideals in the next section.

4. COMPUTING SLOPES OF $\mathcal{H}_{(1,a_2,\dots,a_n)}(\beta)$

Put $A = (1, a_2, \dots, a_n)$, $a_1 = 1 < a_2 < \dots < a_n$. We will evaluate the slopes of the GKZ hypergeometric ideal $H_A(\beta)$ associated to the $1 \times n$ matrix A and $\beta \in \mathbf{C}$ by using the general algorithm given in Section 3. To apply this algorithm, we need to find non-micro-characteristic varieties and compute the restrictions of $\mathcal{H}_A(\beta)$ to these varieties. When f_1, \dots, f_m are polynomials in $\mathbf{C}[x_1, \dots, x_n, \xi_1, \dots, \xi_n]$, we denote by $\mathcal{V}(f_1, \dots, f_m)$ the affine subvariety in \mathbf{C}^{2n} defined by the f_i .

The following theorem can be shown by a standard method of Koszul complex ([1], [6]).

Theorem 4.1. *The characteristic variety of $\mathcal{H}_A(\beta)$ is $\mathcal{V}(\xi_1, \dots, \xi_{n-1}, x_n \xi_n)$. In particular, the singular locus of $\mathcal{H}_A(\beta)$ is $x_n = 0$.*

Note that there is no slope along $x_i = 0$, $1 \leq i \leq n - 1$, which can be shown easily.

Recall that $F = (0, \dots, 0, 1, \dots, 1)$ and $V = (0, \dots, 0, -1, 0, \dots, 0, 1)$. For a positive number p and a non-negative number q , we define the weight vector $L = pF + qV$. In Section 2, we explained the notion of non-micro-characteristic. When the variety is $y_i = x_i = 0$, this notion is rephrased as follows: for a given left

A_n -module A_n/I , the hyperplane $y_i = 0$ ($1 \leq i \leq n-1$) is called non-micro-characteristic of type L when

$$\sqrt{\langle \sigma^L(I), y_i, y_j^*, (j \neq i), \tau^* \rangle} \ni y_i^*.$$

Proposition 4.2. *For the hypergeometric A_n -module $\mathcal{H}_A(\beta)$, the variety $y_i = 0$ ($1 \leq i \leq n-2$) is non-micro-characteristic of type L for all $L = pF + qV$, $p > 0$.*

Proof. Consider $\partial_i^{a_j} - \partial_j^{a_i} \in H_A(\beta)$. For all L and for $i < j \leq n-1$, we have $\sigma^L(\partial_i^{a_j} - \partial_j^{a_i}) = (y_i^*)^{a_j}$, which implies that $y_i = 0$ is non-micro-characteristic of type L . \square

Now, let us apply the second step of the algorithm to evaluate the slopes, i.e., we will compute the restriction of $\mathcal{H}_A(\beta)$ to $y_i = x_i = 0$ ($1 < i \leq n-2$). Let s be an indeterminate. Consider the ideal $H_A[s]$ in $A_n[s]$ generated by $A\theta - s$ and I_A .

Theorem 4.3. *We have a left $D'[s]$ -module isomorphism*

$$(4.1) \quad A_n[s]/(A_n[s]H_A[s] + x_i A_n[s]) \simeq D'[s]/D'[s]H_{A'}[s], \quad i \neq 1.$$

Here, $D' = \mathbf{C}\langle x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n, \partial_1, \dots, \partial_{i-1}, \partial_{i+1}, \dots, \partial_n \rangle$ and $A' = (1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$.

Proof. Fix an order \succ such that $\partial_i \succ \partial_n \succ \partial_{n-1} \succ \dots \succ \partial_1$. Then,

$$\partial_n - \partial_1^{a_n}, \partial_{n-1} - \partial_1^{a_{n-1}}, \dots, \partial_2 - \partial_1^{a_2}$$

is the reduced Gröbner basis of I_A with respect to \succ .

For each i ($2 \leq i \leq n$), by applying the same method as in the proof of [17, Theorem 3.1.3], we can prove that $\text{in}_{(-e_i, e_i)}(H_A(s))$ is generated by $A\theta - s$ and $\text{in}_{e_i}(\partial_j - \partial_1^{a_j})$ ($j = 2, \dots, n$).

Define the V_i -filtration $F_k[s]$ of $A_n[s]$ by

$$F_k[s] = \left\{ \sum a_{\alpha\beta\gamma} x^\alpha \partial^\beta s^\gamma \mid (\beta - \alpha) \cdot e_i \leq k \right\}.$$

We compute the restriction of $A_n[s]/H_A(s)$ to $x_i = 0$ by Oaku's algorithm (see, e.g., [17, Theorem 5.2.6, Algorithm 5.2.8]). Since $\text{in}_{(-e_i, e_i)}(x_i \partial_i - x_i \partial_1^{a_i}) = x_i \partial_i$, the b -function of $H_A(s)$ along $x_i = 0$ is $b(p) = p$. Therefore, by [17, Theorem 5.2.6], we have the following isomorphism of left $D'[s]$ -modules:

$$\begin{aligned} & A_n[s]/(H_A(s) + x_i A_n[s]) \\ & \simeq F_0[s]/\left(F_0[s](A\theta - s) + \sum_{j=2, j \neq i}^n F_0[s](\partial_j - \partial_1^{a_j}) + F_{-1}[s](\partial_i - \partial_1^{a_i}) + x_i F_1[s]\right) \\ & \simeq D'[s]/\left(D'[s](A\theta - s) + \sum_{j=2, j \neq i}^n D'[s](\partial_j - \partial_1^{a_j})\right). \end{aligned}$$

\square

We can specialize s to any complex number β . In fact, we have the following theorem.

Theorem 4.4.

$$A_n/(A_n H_A(\beta) + x_i A_n) \simeq D'/D' H_{A'}(\beta), \quad i \neq 1.$$

Proof. Put $L = x_1\partial_1 + a_2x_2\partial_2 + \cdots + a_nx_n\partial_n - \beta$. We note that $[\partial_k - \partial_1^{a_k}, L] = a_k(\partial_k - \partial_1^{a_k})$. Fix a lexicographic order \prec such that $x_1 \succ \cdots \succ x_n \succ \partial_i \succ \partial_n \succ \partial_{n-1} \succ \cdots \succ \partial_1$. Then, we can show that

$$\{L, \partial_n - \partial_1^{a_n}, \partial_{n-1} - \partial_1^{a_{n-1}}, \dots, \partial_2 - \partial_1^{a_2}\}$$

is a Gröbner basis of $H_A(\beta)$ with respect to the order $\succ_{(-e_i, e_i)}$ by Buchberger's S-pair criterion. In fact, we have

$$\text{sp}(\partial_k - \partial_1^{a_k}, x_1\partial_1 + a_2x_2\partial_2 + \cdots + a_nx_n\partial_n - \beta) = L(\partial_k - \partial_1^{a_k}) - (\partial_k - \partial_1^{a_k})L \rightarrow 0.$$

The remaining checks are easy. Therefore, we conclude that $\text{in}_{(-e_i, e_i)}(H_A(\beta))$ is generated by $A\theta - \beta$ and $\text{in}_{e_i}(\partial_j - \partial_1^{a_j})$ ($j = 2, \dots, n$). The rest of the proof is the same as that of Theorem 4.3. \square

The theorem means that the restriction of $\mathcal{H}_A(\beta)$ to $x_i = 0$ can be exactly expressed in terms of the GKZ system for smaller A . By applying our Algorithm 3.5 for computing slopes by reduction of the number of the variables to the variables x_2, \dots, x_{n-2} , we obtain the following theorem from Proposition 4.2 and Theorem 4.4.

Theorem 4.5. *The geometric slopes of $\mathcal{H}_A(\beta)$ along $x_n = 0$ at the origin and $\mathcal{H}_{(1, a_{n-1}, a_n)}(\beta)$ along $x_3 = 0$ at the origin coincide.*

Example 4.6. We note that (the algebraic slopes) \neq (the geometric slopes) in general. For example, let us apply the ACG algorithm to get the algebraic slopes of $H_A(-30)$ for $A = (1, 3, 7)$. This ideal is generated by

$$x_1\partial_1 + 3x_2\partial_2 + 7x_3\partial_3 + 30, \partial_1^3 - \partial_2, -\partial_1^2\partial_2 + \partial_3, \partial_2^3 - \partial_1^2\partial_3.$$

The output is

$$\{-1, -3/4, -1/2\}.$$

On the other hand, if we apply the ACG algorithm to get the geometric slopes, the output is $\{-3/4\}$.

Example 4.7. (the slopes of $\mathcal{H}_{(1, a_{n-1}, a_n)}(\beta)$) \neq (the slopes of $\mathcal{H}_{(1, a_n)}(\beta)$) in general.

Let us take the example: $A = (1, 3, 7)$. Consider the hypergeometric A_3 -module A_3/I where $I = H_{(1, 3, 7)}(-30)$. As we have seen, the slope of this system along $x_3 = 0$ is $\{-3/4\}$.

Consider $\text{in}_L(I)$ for $L = F + 4V$. By computing the Gröbner basis with respect to L , we can see that

$$\mathcal{V}(\text{in}_L(I)) = \mathcal{V}(x_2, \xi_1, \xi_3) \cup \mathcal{V}(\xi_1, \xi_2, \xi_3).$$

It is not included in $\mathcal{V}(\xi_2)$. Hence $x_2 = 0$ is micro-characteristic of type L , and we cannot apply the restriction criterion.

The condition “non-microcharacteristic for all the filtration $pF + qV$ ” cannot be taken as a way to evaluate the slopes by the restriction. In fact, it can be easily checked by the ACG algorithm that the set of the geometric slopes of $\mathcal{H}_{(1, a_n)}(\beta)$ is equal to $\{1/(1 - a_n)\}$. Hence, the set of the slopes of $\mathcal{H}_{(1, 7)}(\beta)$ is $\{-1/6\}$, which is not equal to $\{-3/4\}$.

We have shown that the computation of the slopes of $\mathcal{H}_A(\beta)$ is reduced to the case of three variables. The slopes in this case are as follows.

Theorem 4.8. (the slopes of $\mathcal{H}_{(1, a_{n-1}, a_n)}(\beta)$) $= \{a_{n-1}/(a_{n-1} - a_n)\}$.

Proof. We fix some notation:

- (1) $A = (1, a, b) \in \mathbf{Z}^3$ and $1 < a < b$.
- (2) $P_1 = \partial_1^a - \partial_2$, $P_2 = \partial_1^b - \partial_3$, $P_3 = \partial_2^b - \partial_3^a$, $P_4 = x_1\partial_1 + ax_2\partial_2 + bx_3\partial_3 - \beta$.
- (3) Let Λ be the linear form with slope $-a/(b-a)$ (i.e., $\Lambda = aF + (b-a)V$).
- (4) Let L, L' be linear forms. We say that $L > L'$ if $\text{slope}(L) > \text{slope}(L')$.
- (5) We will write $y_1 = x_1$, $y_2 = x_2$, $t = x_3$.

The operators P_1, P_2, P_3, P_4 are in $H = H_A(\beta)$. Then, we have the following claims.

- (1) For all linear forms L we have $\sigma^L(P_1) = (\eta_1^*)^a$ and so $\eta_1^* \in \sqrt{\sigma^L(H)}$ for all L .
- (2) For all linear forms L we have $\sigma^L(P_4) = y_1\eta_1^* + ay_2\eta_2^* + b\tau^*(-\tau)$.
- (3) For all linear forms $L > \Lambda$ we have $\sigma^L(P_3) = (\tau)^a$ and so $\tau \in \sqrt{\sigma^L(H)}$ for all $L > \Lambda$.
- (4) Thus, for all $L > \Lambda$ we have $\text{Ch}^L(\mathcal{H}) \subset T_{y_2=0}^*\mathbf{C}^3 \cup T_{\mathbf{C}^3}^*\mathbf{C}^3$, and then $\sqrt{\sigma^L(H)}$ is bihomogeneous and L is not a geometric slope of \mathcal{H} .
- (5) On the other hand, for $L < \Lambda$, we have $\sigma^L(P_3) = (\eta_2^*)^b$. Then $\eta_2^* \in \sqrt{\sigma^L(H)}$ and $\text{Ch}^L(\mathcal{H}) \subset T_{t=0}^*\mathbf{C}^3 \cup T_{\mathbf{C}^3}^*\mathbf{C}^3$. So L is not a geometric slope of \mathcal{H} .
- (6) Therefore, the only possible geometric slope of \mathcal{H} is Λ .

Now, suppose that Λ is not a slope. Then, there is no slope, which implies that the L -characteristic variety $\text{Ch}^L(\mathcal{H}_A(\beta))$ is the same for all $L = pF + qV$, $p, q > 0$ by [3] and [10, Théorème 3.4.1]. It follows from

$$[P_1, P_2] = 0, [P_1, P_4] = aP_1, [P_2, P_4] = bP_2$$

and the Buchberger algorithm that $\{P_1, P_2, P_4\}$ is a Gröbner basis for the order defined by the weight vector $L = (0, 0, -N, 1, 1, N+1)$, $N \geq b$ and a tie-breaking term order such that $x_2 \succ \partial_3 \succ \partial_1 \succ \partial_2$. Therefore, the initial ideal $\text{in}_L(H_A(\beta))$ is generated by ξ_1^a , ξ_3 , $x_1\xi_1 + ax_2\xi_2 + bx_3\xi_3$ and hence the L -characteristic variety is equal to $T_{y_2=0}^*\mathbf{C}^3 \cup T_{\mathbf{C}^3}^*\mathbf{C}^3$. This fact contradicts that the L -characteristic variety is the same for all L . \square

5. RATIONAL SOLUTIONS AND REDUCIBILITY

Our ultimate aim in studying the slopes of $H_A(\beta)$ is to get a better understanding of solutions of this system. We are far from the goal, but to this end, it will be useful to present some facts on classical solutions and a relation to generalized confluent hypergeometric functions.

In the case of the hypergeometric ideal associated to homogeneous monomial curves, Cattani, D'Andrea, and Dickenstein [5] studied rational solutions and reducibility of the system. We will study rational solutions and reducibility of our system.

Theorem 5.1. *Any rational solution of the hypergeometric system $H_{(1,a_2,\dots,a_n)}(\beta)$ is a polynomial. It has a polynomial solution if and only if $\beta \in \mathbf{N} = \{0, 1, \dots\}$. The polynomial solution is the residue of $\exp(\sum x_i t^{a_i}) t^{-\beta}$ at the origin $t = 0$:*

$$\int_C \exp\left(\sum x_i t^{a_i}\right) t^{-\beta} \frac{dt}{t}.$$

Here, C is a circle that encircles the origin in the positive direction.

Proof. Since the singular locus of $H_A(\beta)$ ($A = (a_1, a_2, \dots, a_n)$, $a_1 = 1$) is $x_n = 0$, any rational solution f is a Laurent polynomial with poles on $x_n = 0$. Take a weight vector $w = (0, 1, 1, \dots, 1)$. Then, we have $\text{in}_w(I_A) = \langle \partial_2, \dots, \partial_n \rangle$. The initial term $\text{in}_w(f)$ is annihilated by

$$\sum a_i \theta_i - \beta, \quad \partial_2, \dots, \partial_n.$$

Therefore, $x_1^\beta = \text{in}_w(f)$. This implies $\beta \in \mathbf{N}$, because f has a pole only on $x_n = 0$.

Take a \mathbf{Z} -basis of $\text{Ker}(\mathbf{Z}^n \xrightarrow{A} \mathbf{Z})$ as

$$(-a_2, 1, 0, \dots, 0), (-a_3, 0, 1, 0, \dots, 0), \dots, (-a_n, 0, \dots, 0, 1)$$

to construct series solutions. Since [17, Proposition 3.4.1] holds for non-homogeneous A as well, a formal series solution g of $H_A(\beta)$ satisfying $\text{in}_w(g) = x_1^\beta$ can be uniquely expressed as

$$(5.1) \quad \sum_{m \in \mathbf{N}^{n-1}} \frac{\beta(\beta-1) \cdots (\beta - \sum m_k a_k + 1)}{m_2! \cdots m_n!} \left(\frac{x_2}{x_1^{a_2}} \right)^{m_2} \cdots \left(\frac{x_n}{x_1^{a_n}} \right)^{m_n} x_1^\beta.$$

When $\beta \in \mathbf{N}$, it is a polynomial. The rest of the theorem is easy to show. \square

Let R be the ring of differential operators of n variables with rational function coefficients over $\mathbf{k} = \mathbf{C}$. A left ideal J of R is called *irreducible* when J is a maximal ideal in R . We will study the reducibility of $R \cdot H_A(\beta)$.

We assume that J is zero-dimensional, i.e., $r = \dim_{\mathbf{C}(x)} R/J < +\infty$. Let $V = V(J)$ be the vector space of holomorphic solutions of J on a simply connected open set contained in the non-singular domain of J . It is known that $\dim_{\mathbf{C}} V = r$. Define $I(V)$ by $R \cdot \{\ell \in R \mid \ell \bullet f = 0 \text{ for all } f \in V\}$. If $J \subset I(V)$, $J \neq I(V)$, then we have $\dim_{\mathbf{C}(x)} R/J < \dim_{\mathbf{C}} V = r$ because of the zero-dimensionality of J . Therefore, we have

$$J = I(V(J)).$$

Under this correspondence of ideals and solutions, a zero-dimensional ideal J of R is reducible if and only if there exists a proper subspace W of the solution space of $V(J)$ such that $0 < \dim_{\mathbf{C}(x)} R/I(W) < \dim_{\mathbf{C}(x)} R/J$. In the case of one variable, the reducibility is equivalent to saying that the generator of the ideal can be factored in R .

Theorem 5.2. *The system of differential equations $R \cdot H_A(\beta)$ is reducible if and only if $\beta \in \mathbf{Z}$.*

Proof. Any curve is Cohen-Macaulay. By applying the theorem of Adolphson [1], the holonomic rank of $H_A(\beta)$ is a_n for all β .

Put $M(\beta) = A_n/H_A(\beta)$. Consider the left A_n -morphism

$$(5.2) \quad \partial_1 : M(\beta) \rightarrow M(\beta+1).$$

It has the inverse when $\beta \neq -1$. Therefore, $M(-1) \simeq M(-2) \simeq M(-3) \simeq \cdots$ and $M(0) \simeq M(1) \simeq M(2) \simeq \cdots$.

When $\beta \in \mathbf{N}$, the system admits a polynomial solution; then it is reducible. It is also easy to see that when $\beta \in \mathbf{Z}_{<0}$, the equation is reducible. In fact, consider the left \mathcal{D} -morphism

$$\partial_1 : M(-1) \rightarrow M(0).$$

It induces a morphism to the solutions by

$$f \rightarrow \partial_1 \bullet f.$$

The solution $f = 1$ of $M(0)$ is sent to zero. So the image of ∂_1 gives a proper subspace of solutions in the solution space of $M(-1)$. To find differential equations for the subspace, take all ℓ such that $\ell\partial_1 \in H_A(0)$. Then, $\{\ell\} \subset H_A(-1)$. By the isomorphism (5.2), we conclude that when $\beta \in \mathbf{Z}_{<0}$, the system is reducible.

Let us prove that the system is irreducible when $\beta \notin \mathbf{Z}$ by applying the result of Beukers, Brownawell, and Heckman [4]. For this purpose, we first construct a convergent series solution of $H_A(\beta)$. Take $w = (1, 1, \dots, 1, 0)$. Then the degree of $\text{in}_w(I_A)$ is equal to a_n . Since I_A contains the elements of the form $\partial_i^{a_n} - \partial_n^{a_i}$, the radical of $\text{in}_w(I_A)$ is $\langle \partial_1, \dots, \partial_{n-1} \rangle$. Therefore, the top-dimensional standard pairs have the form $(\partial^b, \{n\})$.

Let v be the zero of the indicial ideal associated to $(\partial^b, \{n\})$:

$$v_1 = b_1, \dots, v_{n-1} = b_{n-1}, v_n = \frac{\beta - \sum a_i b_i}{a_n}.$$

Assume $\beta \notin \mathbf{Z}$ or $\beta \gg 0$. Taking the lattice basis

$$(a_2, -1, 0, \dots, 0), \dots, (a_n, 0, \dots, 0, -1),$$

we have the following a_n linearly independent convergent series solutions:

$$(5.3) \quad \sum_{m \in \mathbf{N}^{n-1}} \frac{v_2(v_2 - 1) \cdots (v_2 - m_2 + 1) \cdots v_n(v_n - 1) \cdots (v_n - m_n + 1)}{(v_1 + 1)(v_1 + 2) \cdots (v_1 + \sum a_k m_k)} \cdot \left(\frac{x_1^{a_2}}{x_2} \right)^{m_2} \cdots \left(\frac{x_1^{a_n}}{x_n} \right)^{m_n} x^v.$$

We consider the change of variables:

$$y_1 = x_1, y_2 = x_1^{a_2}/x_2, \dots, y_n = x_1^{a_n}/x_n.$$

The inverse of this change of variables is also rational, and the change of variables induces that of ∂_i . We denote by Φ the operation of this change of variables of x_i and ∂_i . Since the irreducibility is invariant under any birational change of variables, we will prove the irreducibility of the ideal $J = R \cdot \Phi(x^{-v} H_A(\beta) x^v)$ where $R = \mathbf{C}(y) \langle \partial_{y_1}, \dots, \partial_{y_n} \rangle$.

Let V be the solution space of J spanned by the series $\Phi((5.3) \cdot x^{-v})$ near $y = 0$. If $J = I(V)$ is reducible, then there exists a proper subspace W of J such that $0 < \dim_{\mathbf{C}(y)} R/I(W) < a_n$. We consider the vector space $W' = \{f(0, \dots, 0, y_n) \mid f \in W\}$. It is easy to see that $\dim_{\mathbf{C}(y_n)} R'/I(W') \leq \dim_{\mathbf{C}} W$ where $R' = \mathbf{C}(y_n) \langle \partial_{y_n} \rangle$. Let us prove that $\dim_{\mathbf{C}} W' = a_n$ when $\beta \notin \mathbf{Z}$, which implies the irreducibility of J by a contradiction.

We restrict the series $\Phi((5.3) \cdot x^{-v})$ to $y_1 = x_1 = 0, y_2 = x_1^{a_2}/x_2 = 0, \dots, y_{n-1} = x_1^{a_{n-1}}/x_n = 0$ and replace $y_n = x_1^{a_n}/x_n$ by z . Without loss of generality, we may assume $v_1 = 0$. Then the restricted series has the form

$$(5.4) \quad \sum_{m=0}^{\infty} \frac{v_n(v_n - 1) \cdots (v_n - m + 1)}{(a_n m)!} (-z)^m.$$

It is annihilated by the ordinary differential operator

$$(a_n\theta_z)(a_n\theta_z - 1) \cdots (a_n\theta_z - a_n + 1) - z(\theta_z + v_n).$$

By replacing $z/a_n^{\alpha_n}$ by x , we obtain the generalized hypergeometric ordinary differential equation

$$(5.5) \quad \theta_x(\theta_x - 1/a_n) \cdots (\theta_x - (a_n - 1)/a_n) - x(\theta_x + v_n).$$

By [4], this ordinary differential equation of rank a_n is reducible if and only if $v_n - k/a_n \notin \mathbf{Z}$ for all $k = 0, 1, \dots, a_n - 1$. If one of them is an integer, β becomes an integer. Therefore, the ideal $I(W')$ contains the principal ideal generated by (5.5), which is maximal when $\beta \notin \mathbf{Z}$. We conclude that $I(W')$ is generated by (5.5) and hence $\dim_{\mathbf{C}} W' = a_n$. \square

ACKNOWLEDGEMENT

The main part of this paper was obtained from discussions during the second author's visit to Universidad de Sevilla in July, 2000. The authors are grateful to the University for providing them the opportunity of intensive discussions.

REFERENCES

- Adolphson, A., Hypergeometric functions and rings generated by monomials, *Duke Mathematical Journal* 73 (1994), 269–290. MR **96c**:33020
- Assi, A., Castro-Jiménez, F. J., and Granger, J. M., How to calculate the slopes of a \mathcal{D} -module, *Compositio Math.* 104 (1996), no. 2, 107–123. MR **98i**:32010
- Assi, A., Castro-Jiménez, F. J., and Granger, M., The Gröbner fan of an A_n -module, *J. Pure Appl. Algebra* 150 (2000), 1, 27–39. MR **2001j**:16036
- Beukers, F., Brownawell, W. D., and Heckman, G., Siegel normality, *Ann. of Math.* (2) 127 (1988), no. 2, 279–308. MR **90e**:11106
- Cattani, E., D'Andrea, C., and Dickenstein, A., The \mathcal{A} -hypergeometric system associated with a monomial curve, *Duke Mathematical J.* 99 (1999), 179–207. MR **2001f**:33018
- Gel'fand, I. M., Zelevinskii, A. V., and Kapranov, M. M., Hypergeometric functions and toric varieties, *Funktsional. Anal. i Prilozhen.*, 23 (1989), 2, 12–26; *English transl.*, *Funct. Anal. Appl.* 23 (1989), no. 2, 94–106. MR **90m**:22025
- Grayson, D. and Stillman, M., Macaulay2, a software system for research in algebraic geometry, available at <http://www.math.uiuc.edu/Macaulay2>.
- Hotta, R., Equivariant D -modules. Preprint math.RT/9805021.
- Laurent, Y. Théorie de la deuxième microlocalisation dans le domaine complexe, *Progress in Math.* 53, Birkhäuser, 1985. MR **86k**:58113
- Laurent, Y., Polygone de Newton et b -fonctions pour les modules microdifférentiels, *Ann. Sci. École Norm. Sup.* (4) 20 (1987), no. 3, 391–441. MR **89k**:58282
- Laurent, Y. and Mebkhout, Z., Pentés algébriques et pentés analytiques d'un \mathcal{D} -module, *Ann. Sci. École Norm. Sup.* (4) 32 (1999), no. 1, 39–69. MR **2001b**:32015
- Laurent, Y. and Mebkhout, Z., Image inverse d'un \mathcal{D} -module et polygone de Newton, *Compositio Math.* 131 (2002), no. 1, 97–119.
- Leykin, A. and Tsai, H., D -module package for Macaulay 2. <http://www.math.cornell.edu/~htsai>
- Mebkhout, Z., Le théorème de positivité de l'irrégularité pour les \mathcal{D}_X -modules, *The Grothendieck Festschrift, Vol. III*, 83–132, *Progr. Math.*, 88, Birkhäuser, Boston, MA, 1990. MR **92j**:32031
- Oaku, T., Algorithms for b -functions, restriction and algebraic local cohomology groups of D -modules, *Advances in Applied Mathematics* 19 (1997), 61–105. MR **98d**:14031
- Oaku, T., Takayama, N., and Walther, U., A localization algorithm for D -modules, *Journal of Symbolic Computation* 29 (2000), 721–728. MR **2001g**:13056

17. Saito, M., Sturmfels, B., and Takayama, N., *Gröbner deformations of hypergeometric differential equations*, Algorithms and Computation in Mathematics, 6. Springer-Verlag, Berlin, 2000. MR **2001i**:13036
18. Takayama, N., *Kan: A system for computation in algebraic analysis*, 1991 version 1, 1994 version 2, the latest version is 3.000726. Source code available for Unix computers. Download from <http://www.openxm.org>

UNIVERSIDAD DE SEVILLA, DEPTO. DE ÁLGEBRA, APDO. 1160, E-41080 SEVILLA, SPAIN
E-mail address: `castro@us.es`

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, KOBE UNIVERSITY, 1-1, ROKKODAI,
NADA-KU, KOBE 657-8501, JAPAN
E-mail address: `takayama@math.kobe-u.ac.jp`

ASYMPTOTICS FOR LOGICAL LIMIT LAWS: WHEN THE GROWTH OF THE COMPONENTS IS IN AN RT CLASS

JASON P. BELL AND STANLEY N. BURRIS

ABSTRACT. Compton's method of proving monadic second-order limit laws is based on analyzing the generating function of a class of finite structures. For applications of his deeper results we previously relied on asymptotics obtained using Cauchy's integral formula. In this paper we develop elementary techniques, based on a Tauberian theorem of Schur, that significantly extend the classes of structures for which we know that Compton's theory can be applied.

1. INTRODUCTION

We are primarily interested in being able to show that an answer exists to the following question: given a class \mathcal{K} of finite relational structures and a property Φ :

What is the probability that a finite structure, randomly selected from \mathcal{K} , has the property Φ ?

While pursuing this goal we are able to prove (in Corollary 4.3) the conjecture ([10], p. 462) of Durrett, Granovsky and Gueron arising in the study of coagulation-fragmentation processes which says that if $\mathbf{S}(x) = \sum s(n)x^n$ is a power series with positive coefficients such that $s(n-1)/s(n) \rightarrow \rho$, then the coefficients of the power series expansion of $\exp(\mathbf{S}(x))$ have the same property.

The notion of probability that we use is to take the proportion q_n of finite structures of size n in \mathcal{K} (we only count up to isomorphism) that have the property Φ , and then to take the limit q of the sequence q_n as n goes to infinity. This limit, when it exists, is also called the *asymptotic density* of the class of members of \mathcal{K} that satisfy Φ .

It has been known since the mid 1970s that a few well-known classes, like graphs or directed graphs, are such that if Φ is a property defined by a first-order sentence, then the probability must exist, and be 0 or 1. If \mathcal{K} is a class such that every sentence in a given (logic) language L defines a property for which a probability exists, then we say \mathcal{K} has an *L-limit law*.

In the 1980s Kevin Compton gave a new method for proving that a class \mathcal{K} of relational structures has a monadic second-order limit law, a method that only depends on analyzing the growth rate of $a(n)$, the number of structures of size n in \mathcal{K} . This applies to classes that are closed under disjoint union and components.

Received by the editors June 26, 2002 and, in revised form, January 10, 2003.

2000 *Mathematics Subject Classification*. Primary 03C13, 05A16, 11P99, 41A60; Secondary 11N45, 11N80, 11U99, 60J20.

Key words and phrases. Ratio test, Schur's Tauberian theorem, asymptotic density, monadic second-order logic, zero-one law, limit law.

The second author would like to thank NSERC for support of this research.

For such classes the count function $p(n)$, the number of components of size n in \mathcal{K} , is often more readily available than $a(n)$, and we would like to know conditions on $p(n)$ that guarantee a logical limit law. The best results of this type previously known were

- if $p(n) = O(n^c)$, that is, $p(n)$ is polynomially bounded, then \mathcal{K} has a monadic second-order 0–1 law (Bell [1]), and
- if $p(n) = C\beta^n + O(\gamma^n)$, where $0 < \gamma < \beta$ and $C > 0$, then \mathcal{K} has a monadic second-order limit law (see [4], Chapters 5 and 6).

We will make use of a Tauberian theorem of Schur to extend these results to cover a large number of new classes of structures. For example, we will show that

$$p(n) \sim an^be^{cn^d},$$

with $d < 1$ and $a, c > 0$, yields a monadic second-order 0–1 law; and

Corollary 9.4. For $\mu < 1 < \beta$ and $C > 0$, or $\mu = 1 < \beta$ and $C > 1$, the condition

$$p(n) \sim C\beta^n/n^\mu$$

guarantees that \mathcal{K} has a monadic second-order limit law.

In addition to the standard “big O” and “little o” notation from asymptotics we use the following:

notation	means
$f(n) \prec g(n)$	$f(n)$ is eventually less than $g(n)$
$f(n) \preceq g(n)$	$f(n)$ is eventually less than or equal to $g(n)$
$f(n) \succ g(n)$	$f(n)$ is eventually greater than $g(n)$
$f(n) \succeq g(n)$	$f(n)$ is eventually greater than or equal to $g(n)$

As mathematics symbols we will be using upper case boldface roman letters exclusively to denote power series, and the corresponding lower case ordinary italic letters name the coefficients. Thus we will use $\mathbf{A}(x) = \sum_n a(n)x^n, \dots, \mathbf{T}(x) = \sum_n t(n)x^n$.

2. THE CLASS RT_ρ

The sequences in RT_ρ play a central role in the study of Compton’s development of logical limit laws. We find conditions to guarantee membership in this class and then apply them to prove logical limit laws.

Definition 2.1. RT_ρ is the collection of sequences $s(n)$ of real numbers that satisfy

- (a) $s(n) \succ 0$, and
- (b) $\lim_{n \rightarrow \infty} \frac{s(n-1)}{s(n)} = \rho$.

We also say that a power series $\mathbf{S}(x)$ is in RT_ρ if its sequence of coefficients $s(n)$ is in RT_ρ . Also, if $f(x)$ is a function that admits a power series expansion $\sum s(n)x^n$ about 0, then we say $f(x)$ is in RT_ρ if $s(n)$ is in RT_ρ .

The following lemma and corollary give the most basic information about the growth rate of members of RT_ρ .

Lemma 2.2. If $s(n) \in \text{RT}_\rho$ with $0 < \rho < \infty$ then, for $0 < \varepsilon < \rho^{-1}$, there is an N such that, for $n \geq N$,

$$(\rho^{-1} - \varepsilon)^n < s(n) < (\rho^{-1} + \varepsilon)^n.$$

From this we see that a smaller ρ leads to much faster-growing sequences.

Corollary 2.3. *If $s(n) \in \text{RT}_\sigma$ and $t(n) \in \text{RT}_\tau$ with $0 < \sigma < \tau < \infty$, then $t(n) = o(s(n))$.*

The class RT_ρ has some remarkable similarities to the class RV_α , the class of functions of regular variation at infinity with index α , but this connection does not seem to have been thoroughly researched. For $0 < \rho < \infty$ the sequence ρ^{-n} is perhaps the simplest member of the class RT_ρ , and this sequence, along with RT_1 , completely determines RT_ρ , since one can easily check that

$$s(n) \in \text{RT}_\rho \iff s(n)\rho^n \in \text{RT}_1.$$

In terms of power series this would be written as

$$(2.1) \quad \mathbf{S}(x) \in \text{RT}_\rho \iff \mathbf{S}(\rho x) \in \text{RT}_1.$$

Here are three of the simplest examples from RT_1 :

$$\begin{aligned} s(n) &= c && \text{for } c > 0, \\ s(n) &= n^c && \text{for } c \text{ any real number,} \\ s(n) &= d^{n^c} && \text{for } c < 1 < d. \end{aligned}$$

We can use these examples, with Proposition 2.4, to make a substantial collection of sequences in RT_ρ . Multiplication of the sequences $s(n)$ and $t(n)$ gives $s(n) \cdot t(n)$, and division gives $s(n)/t(n)$, where we define $s(n)/t(n)$ to be 0 whenever $t(n) = 0$.

Proposition 2.4. *RT_1 is closed under multiplication, division, and asymptotically equal. For $0 < \sigma, \tau < \infty$, if $s(n) \in \text{RT}_\sigma$ and $t(n) \in \text{RT}_\tau$, then $1/s(n) \in \text{RT}_{1/\sigma}$, and $s(n) \cdot t(n) \in \text{RT}_{\sigma\tau}$. Furthermore, $\mathbf{S}(x) \in \text{RT}_\rho$ iff $\mathbf{S}'(x) \in \text{RT}_\rho$ iff $x\mathbf{S}(x) \in \text{RT}_\rho$.*

Proof. (Straightforward) □

With this we can easily see, for example, that for $a, c, \rho > 0$ and b any real number, if $s(n) \sim an^be^{c\sqrt{n}}/\rho^n$, then $s(n) \in \text{RT}_\rho$. Just such an example played an important role in finding the first applications of Compton's 1989 theoretical development of logical limit laws. A key fact about this particular sequence is that $an^be^{c\sqrt{n}} \in \text{RT}_1$, and it is eventually nondecreasing.

3. THE CAUCHY PRODUCT

The Cauchy product $\mathbf{R}(x) = \mathbf{S}(x) \cdot \mathbf{T}(x)$ is defined by

$$r(n) = \sum_{k=0}^n s(k) \cdot t(n-k).$$

The following two lemmas and corollary help us extract information about RT_ρ classes from the Cauchy product. We start with the classic Tauberian theorem of Schur.¹

Lemma 3.1 (Schur). *With $0 \leq \rho < \infty$, suppose that*

$$(a) \quad \mathbf{A}(x) \in \text{RT}_\rho,$$

¹It is a Tauberian theorem since it gives an extra condition, namely the radius of convergence of $\mathbf{B}(x) > \rho$, to allow us to go from a generalized limit, $\lim_{x \rightarrow \rho} \mathbf{C}(x)/\mathbf{A}(x)$, to an ordinary limit,

$\lim_{n \rightarrow \infty} c(n)/a(n)$.

- (b) $\mathbf{B}(x)$ has radius of convergence greater than ρ , and
 (c) $\mathbf{B}(\rho) > 0$.

Let $\mathbf{C}(x) = \mathbf{A}(x) \cdot \mathbf{B}(x)$. Then

$$c(n) \sim \mathbf{B}(\rho) \cdot a(n).$$

Proof. (See Bender [3] or Burris [4].) □

Thus from the hypotheses of Schur's Lemma we deduce that $\mathbf{C}(x) \in \text{RT}_\rho$.

Corollary 3.2. *With $0 < \rho < \infty$, suppose $\mathbf{A}(x) \in \text{RT}_\rho$ and the radius of convergence of $\mathbf{B}(x)$ is greater than ρ . If $\mathbf{B}(\rho) > 0$, then*

$$\mathbf{A}(x) \cdot \mathbf{B}(x) \in \text{RT}_\rho.$$

Proof. Let $\mathbf{C}(x) = \mathbf{A}(x) \cdot \mathbf{B}(x)$. From Schur's Lemma we have

$$c(n) \sim \mathbf{B}(\rho) \cdot a(n);$$

so, by Proposition 2.4, $\mathbf{C}(x) \in \text{RT}_\rho$. □

The next lemma offers a variation on this theme.

Lemma 3.3. *Suppose $\mathbf{C}(x) = \mathbf{A}(x) \cdot \mathbf{B}(x)$, where $\mathbf{A}(x)$ and $\mathbf{B}(x)$ have nonnegative coefficients and $\mathbf{B}(x)$ is not the zero power series. If*

- (a) $\mathbf{A}(x) \in \text{RT}_\rho$, where $0 < \rho < \infty$,
 (b) $b(n) = o(c(n))$, and
 (c) $\frac{c(n-1)}{c(n)} \preccurlyeq \rho$,

then $\mathbf{C}(x) \in \text{RT}_\rho$.

Proof. From the following equivalent statements,

$$\begin{array}{ll} \mathbf{A}(x) = \mathbf{B}(x) \cdot \mathbf{C}(x) & \text{iff } \mathbf{A}(\rho x) = \mathbf{B}(\rho x) \cdot \mathbf{C}(\rho x), \\ \mathbf{A}(x) \in \text{RT}_\rho & \text{iff } \mathbf{A}(\rho x) \in \text{RT}_1, \\ b(n) = o(c(n)) & \text{iff } b(n)\rho^n = o(c(n)\rho^n), \\ \frac{c(n-1)}{c(n)} \preccurlyeq \rho & \text{iff } \frac{c(n-1)\rho^{n-1}}{c(n)\rho^n} \preccurlyeq 1, \\ \mathbf{C}(x) \in \text{RT}_\rho & \text{iff } \mathbf{C}(\rho x) \in \text{RT}_1, \end{array}$$

it suffices to prove the lemma in the case that $\rho = 1$. With $\rho = 1$ the goal is:

from (a') $\mathbf{A}(x) \in \text{RT}_1$, (b') $b(n) = o(c(n))$, and (c') $c(n-1) \preccurlyeq c(n)$,
 prove that $\mathbf{C}(x) \in \text{RT}_1$.

Let $\varepsilon > 0$. By (a') there exists an integer M such that, for $n > M$,

$$(3.1) \quad |a(n) - a(n-1)| < \varepsilon a(n).$$

Since $\mathbf{A}(x) \in \text{RT}_1$ implies $a(n) \succ 0$, and $\mathbf{B}(x)$ is not the zero power series, we have $c(n) \succ 0$. So, in view of (c') we can also assume that M is sufficiently large to guarantee

$$(3.2) \quad n \geq M \implies \begin{cases} c(n) > 0, & \text{and} \\ \frac{c(n-1)}{c(n)} \leq 1. \end{cases}$$

By assumption (b') we can find an integer $N > M$ such that

$$(3.3) \quad b(n) \leq \frac{\varepsilon}{(M+1) \max(1, a(0), \dots, a(M))} \cdot c(n)$$

for $n \geq N - M$.

Now suppose $n \geq M + N$. Then, by (3.2),

$$(3.4) \quad 0 < c(n-M) \leq \dots \leq c(n-1) \leq c(n).$$

For $n \geq M + N$,

$$\begin{aligned} c(n) - c(n-1) &= \sum_{i=0}^n a(n-i) \cdot b(i) - \sum_{i=0}^{n-1} a(n-1-i) \cdot b(i) \\ &= \sum_{i < n-M} (a(n-i) - a(n-1-i)) \cdot b(i) \\ &\quad + \sum_{i=n-M}^n a(n-i) \cdot b(i) - \sum_{i=n-M}^{n-1} a(n-1-i) \cdot b(i) \\ &\leq \varepsilon \sum_{i=0}^n a(n-i) \cdot b(i) + \sum_{i=n-M}^n \varepsilon c(i)/(M+1) \quad \text{by (3.1), (3.3)} \\ &\leq \varepsilon c(n) + \sum_{i=n-M}^n \varepsilon c(n)/(M+1) \quad \text{by (3.4)} \\ &= 2\varepsilon c(n). \end{aligned}$$

Therefore, $1 - 2\varepsilon \leq \frac{c(n-1)}{c(n)} \leq 1$. Since this holds for any $n > M + N$, we have

$$\lim_{n \rightarrow \infty} \frac{c(n-1)}{c(n)} = 1, \text{ and so } \mathbf{C}(x) \in \mathbf{RT}_1. \quad \square$$

4. EXPONENTIATION

We will be particularly interested in knowing that exponentiation of a power series preserves membership in \mathbf{RT}_ρ . First we prove a special case of this result.

Lemma 4.1. *Let $\mathbf{T}(x) = \exp(\mathbf{S}(x))$, where*

- (a) $\mathbf{S}(x) \in \mathbf{RT}_1$,
- (b) *the $s(n)$ are nonnegative, and*
- (c) $\liminf_{n \rightarrow \infty} \frac{t(n)}{t(n-1)} = C > 0$.

Then $\mathbf{T}(x) \in \mathbf{RT}_1$.

Proof. Note that $C \leq 1$ since the radius of convergence of $\mathbf{T}(x)$ is 1. From $\mathbf{S}(x) \in \mathbf{RT}_1$ we know that $x\mathbf{S}'(x) \in \mathbf{RT}_1$. So, given $\varepsilon > 0$, we can choose M such that

$$|ms(m) - (m-1)s(m-1)| < \varepsilon ms(m)$$

for $m \geq M$. Also, we can choose N such that $\frac{t(n)}{t(n-1)} > \frac{C}{2}$ for $n > N$. From this we see that

$$(4.1) \quad n-r > N \implies \frac{t(n)}{t(n-r)} > \frac{C^r}{2^r}.$$

Differentiating $\mathbf{T}(x) = \exp(\mathbf{S}(x))$ we have

$$\mathbf{T}'(x) = \mathbf{S}'(x)\mathbf{T}(x),$$

and equating the coefficients of x^{n-1} on both sides of this equation gives

$$(4.2) \quad nt(n) = \sum_{m \leq n} t(m) \cdot (n-m)s(n-m).$$

From this it follows that if $n > N + M$ (we adopt the convention that $s(m) = 0$ for $m < 0$),

$$\begin{aligned} & |nt(n) - (n-1)t(n-1)| \\ &= \left| \sum_{m \leq n} t(m) \cdot ((n-m)s(n-m) - (n-1-m)s(n-1-m)) \right| \\ &\leq \sum_{0 \leq m \leq n-M} t(m) \cdot \varepsilon(n-m)s(n-m) \\ &\quad + \sum_{n-M < m \leq n} t(m) \cdot |(n-m)s(n-m) - (n-1-m)s(n-1-m)| \\ &\leq \varepsilon nt(n) + 2 \cdot \left(\max_{0 \leq m \leq M} ms(m) \right) \cdot \sum_{n-M < m \leq n} t(m) \quad \text{by (4.2)} \\ &\leq \varepsilon nt(n) + 2M \cdot \left(\max_{0 \leq m \leq M} s(m) \right) \cdot M2^M \cdot C^{-M} \cdot t(n) \quad \text{by (4.1)} \\ &= (\varepsilon + o(1))nt(n). \end{aligned}$$

Thus $nt(n) \in \text{RT}_1$; so $t(n) \in \text{RT}_1$. □

Lemma 4.2. $\mathbf{S}(x) \in \text{RT}_1 \implies \exp(\mathbf{S}(x)) \in \text{RT}_1$.

Proof. Let $\mathbf{T}(x) = \exp(\mathbf{S}(x))$. Since $s(n) \in \text{RT}_1$ and $2^{-n} \in \text{RT}_2$, by Corollary 2.3 there exists N such that $s(n) > 2^{-n}$ for $n > N$. Define $\widehat{s}(n)$ to be 1 for $0 \leq n \leq N$ and to be $s(n)$ for $n > N$. Let $\widehat{\mathbf{S}}(x) = \sum_n \widehat{s}(n)x^n$, and let $\widehat{\mathbf{T}}(x) = \sum_n \widehat{t}(n)x^n$ be such that $\widehat{\mathbf{T}}(x) = \exp(\widehat{\mathbf{S}}(x))$. Since

$$\widehat{s}(n) \geq 2^{-n}/n, \quad \text{for } n \geq 1,$$

$\widehat{\mathbf{S}}(x) + \log(1-x/2)$ is a power series with nonnegative coefficients. Thus, for $n \geq 0$,

$$\begin{aligned} \widehat{t}(n) - \widehat{t}(n-1)/2 &= [x^n](1-x/2)\widehat{\mathbf{T}}(x) \\ &= [x^n]\exp(\log(1-x/2) + \widehat{\mathbf{S}}(x)) \\ &\geq 0; \end{aligned}$$

so

$$\liminf_{n \rightarrow \infty} \frac{\widehat{t}(n)}{\widehat{t}(n-1)} \geq 1/2 > 0.$$

Now $\widehat{\mathbf{T}}(x) \in \text{RT}_1$ by Lemma 4.1 since $\widehat{\mathbf{S}}(x) \in \text{RT}_1$. To finish the proof, observe that

$$\mathbf{S}(x) - \widehat{\mathbf{S}}(x) = \mathbf{p}(x),$$

where $\mathbf{p}(x)$ is a polynomial; so

$$\mathbf{T}(x) = \exp(\mathbf{p}(x)) \cdot \widehat{\mathbf{T}}(x).$$

By Corollary 3.2, $\mathbf{T}(x) \in \text{RT}_1$. □

Next we see that RT_ρ classes are closed under exponentiation of power series, proving the conjecture ([10], p. 462) of Durrett, Granovsky and Gueron.

Corollary 4.3. For $0 < \rho < \infty$,

$$\mathbf{S}(x) \in \text{RT}_\rho \implies \exp(\mathbf{S}(x)) \in \text{RT}_\rho.$$

Proof. This follows from (2.1) and Lemma 4.2. \square

If we exponentiate a power series in RT_ρ that diverges at ρ , then we obtain a power series whose coefficients grow much faster than those of the original series.

Lemma 4.4. Let $\mathbf{T}(x) = \exp(\mathbf{S}(x))$ where

- (a) $\mathbf{S}(x) \in \text{RT}_\rho$ with $0 < \rho < \infty$, and
- (b) $\mathbf{S}(\rho) = \infty$.

Then $s(n) = o(t(n))$.

Proof. First we prove this lemma for the case that the coefficients of $\mathbf{S}(x)$ are nonnegative. Choose $N \geq 1$ and observe that, for $n \geq N$,

$$t(n) = \sum_{j \geq 0} \frac{1}{j!} [x^n] \mathbf{S}(x)^j \geq \frac{1}{2} [x^n] \mathbf{S}(x)^2 \geq \frac{1}{2} \sum_{k=0}^N s(k) s(n-k)$$

and thus

$$\frac{t(n)}{s(n)} \geq \frac{1}{2} \sum_{k=0}^N s(k) \frac{s(n-k)}{s(n)}.$$

Taking the $\liminf_{n \rightarrow \infty}$ of both sides, using the fact that $\mathbf{S}(x) \in \text{RT}_\rho$, gives

$$\liminf_{n \rightarrow \infty} \frac{t(n)}{s(n)} \geq \frac{1}{2} \sum_{k=0}^N s(k) \rho^k.$$

Now use the fact that $\mathbf{S}(\rho) = \infty$.

For the general case, where some of the coefficients of $\mathbf{S}(x)$ can be negative, let $\mathbf{p}(x)$ be a polynomial such that $\widehat{\mathbf{S}}(x) = \mathbf{S}(x) + \mathbf{p}(x)$ has nonnegative coefficients. Clearly $\widehat{\mathbf{S}}(x) \in \text{RT}_\rho$ and $\widehat{\mathbf{S}}(\rho) = \infty$. Then $\widehat{\mathbf{T}}(x) = \exp(\widehat{\mathbf{S}}(x))$ is such that

$$(4.3) \quad \widehat{s}(n) = o(\widehat{t}(n)),$$

by the first part of the proof. By Corollary 4.3, $\widehat{\mathbf{T}}(x) \in \text{RT}_\rho$. Since

$$\mathbf{T}(x) = \exp(-\mathbf{p}(x)) \cdot \widehat{\mathbf{T}}(x),$$

by Schur's Lemma we have

$$t(n) \sim \exp(-\mathbf{p}(\rho)) \cdot \widehat{t}(n).$$

So from (4.3) we have $s(n) = o(t(n))$ since $s(n)$ eventually equals $\widehat{s}(n)$. \square

A modest growth condition on the coefficients of $\mathbf{S}(x)$ guarantees that the coefficients of $\exp(\mathbf{S}(x))$ satisfy a growth condition used to prove logical limit laws. But first we prove a technical lemma on membership in RT_1 .

Lemma 4.5. Suppose $\mathbf{S}(x) \in \text{RT}_1$. Then

$$\liminf_{n \rightarrow \infty} s(n) > 1 \implies s(n) - 1 \in \text{RT}_1.$$

Proof. Choose $C, N > 1$ such that $s(n) \geq C$, for $n \geq N$. Then, for $n \geq N$,

$$(4.4) \quad 1 - \frac{1}{s(n)} \geq \frac{C-1}{s(n)} \geq C-1 > 0.$$

Now

$$\frac{s(n-1)-1}{s(n)-1} - 1 = \frac{\frac{s(n-1)}{s(n)} - 1}{1 - \frac{1}{s(n)}},$$

and the right side tends to 0 as n tends to infinity since $s(n) \in \text{RT}_1$ says the numerator tends to 0, and (4.4) says the denominator is bounded away from 0. Thus,

$$\lim_{n \rightarrow \infty} \frac{s(n-1)-1}{s(n)-1} = 1.$$

□

Now we proceed with the analysis of the growth rate of the coefficients after exponentiation.

Lemma 4.6. *Let $\mathbf{T}(x) = \exp(\mathbf{S}(x))$ where*

- (a) $\mathbf{S}(x) \in \text{RT}_\rho$ with $0 < \rho < \infty$, and
- (b) $\liminf_{n \rightarrow \infty} ns(n)\rho^n > 1$.

Then $\mathbf{T}(x) \in \text{RT}_\rho$ and $\frac{t(n-1)}{t(n)} \prec \rho$.

Proof. Corollary 4.3 gives $\mathbf{T}(x) \in \text{RT}_\rho$. To verify $t(n-1) \prec t(n)\rho$ note that

$$\begin{aligned} t(n)\rho^n - t(n-1)\rho^{n-1} &= [x^n] \left((1-x) \cdot \mathbf{T}(\rho x) \right) \\ &= [x^n] \left((1-x) \cdot \exp(\mathbf{S}(\rho x)) \right) \\ &= [x^n] \exp \left(s(0) + \sum_{n \geq 1} (s(n)\rho^n - 1/n)x^n \right). \end{aligned}$$

Since $\mathbf{S}(x) \in \text{RT}_\rho$,

$$ns(n)\rho^n \in \text{RT}_1,$$

and then Lemma 4.5 gives $ns(n)\rho^n - 1 \in \text{RT}_1$. Using Proposition 2.4 we have

$$(4.5) \quad s(n)\rho^n - 1/n \in \text{RT}_1.$$

Choose $N \geq 1$ such that, for $n \geq N$,

$$(4.6) \quad ns(n)\rho^n > 1,$$

and let

$$\begin{aligned} \mathbf{P}(x) &= s(0) + \sum_{n=1}^{N-1} (s(n)\rho^n - 1/n)x^n - \sum_{n=1}^{N-1} x^n, \\ \mathbf{R}(x) &= \sum_{n=1}^{N-1} x^n + \sum_{n=N}^{\infty} (s(n)\rho^n - 1/n)x^n. \end{aligned}$$

By (4.5) we know that $\mathbf{R}(x) \in \text{RT}_1$; so Lemma 4.2 gives

$$(4.7) \quad \exp(\mathbf{R}(x)) \in \text{RT}_1.$$

Noting that $\mathbf{p}(x)$ is a polynomial we have

$$\begin{aligned} t(n)\rho^n - t(n-1)\rho^{n-1} &= [x^n] \exp\left(s(0) + \sum_{n \geq 1} (s(n)\rho^n - 1/n)\right) \\ &= [x^n] (\exp(\mathbf{p}(x)) \cdot \exp(\mathbf{R}(x))) \\ &\sim \exp(\mathbf{p}(1)) \cdot [x^n] \exp(\mathbf{R}(x)) \end{aligned}$$

by (4.7) and Schur's Lemma, and thus, since the coefficients of $\mathbf{R}(x)$ are positive,

$$t(n)\rho^n - t(n-1)\rho^{n-1} \succ 0.$$

This says $\frac{t(n-1)}{t(n)} \prec \rho$. □

5. THE STAR TRANSFORMATION

Now we introduce a transformation on a power series $\mathbf{S}(x)$ that plays an important role in combinatorics.

Definition 5.1. Let $\mathbf{S}^*(x) = \sum s^*(n)x^n$ be the power series defined by

$$\begin{aligned} s^*(0) &= 0, \\ s^*(n) &= \sum_{j+k=n} s(j)/k \quad \text{for } n \geq 1. \end{aligned}$$

Lemma 5.2. For $0 < \rho < 1$, if $\mathbf{S}(x) \in \text{RT}_\rho$ has nonnegative coefficients, then

- (a) $s^*(n) \sim s(n)$, and
- (b) $\mathbf{S}^*(x) \in \text{RT}_\rho$.

Proof. Since $s(n) \in \text{RT}_\rho$ we know from Corollary 2.3 that $0 < \beta < \rho^{-1} < \alpha$ implies

$$n^2\beta^n = o(s(n)) \quad \text{and} \quad s(n) = o(\alpha^n)$$

since $n^2\beta^n \in \text{RT}_{1/\beta}$ and $\alpha^n \in \text{RT}_{1/\alpha}$. Choose α satisfying $\rho^{-1} < \alpha < \rho^{-2}$ and choose C such that $|s(n)| < C\alpha^n$ for all n . Choose N such that $s(n) \geq 0$ for $n \geq N$. Then

$$\begin{aligned} ns^*(n) &= \sum_{d|n} ds(d) \\ &\leq ns(n) + \sum_{d \leq n/2} ds(d) \\ &\leq ns(n) + \sum_{d \leq n/2} dC\alpha^{n/2} \\ &= ns(n) + O(n^2\alpha^{n/2}) \\ &= ns(n) + o(s(n)) \quad (\text{since } \sqrt{\alpha} < \rho^{-1}). \end{aligned}$$

Hence $ns^*(n) \sim ns(n)$ since $0 \leq s(n) \leq s^*(n)$, and thus $s^*(n) \sim s(n)$, giving (a). Then from Proposition 2.4 we have $\mathbf{S}^*(x) \in \text{RT}_\rho$. □

This result is best possible since one can easily find examples with $\rho = 1$ such that $\mathbf{S}(x) \in \text{RT}_1$ but $\mathbf{S}^*(x) \notin \text{RT}_1$. For example, $\mathbf{S}(x) = \sum_n x^n \in \text{RT}_1$, but $\mathbf{S}^*(x) = \sum_n \sigma(n)x^n/n \notin \text{RT}_1$. However, the power series expansion of $\frac{x}{1-x} \cdot \mathbf{S}'(x)$ is much better behaved in this situation.

Lemma 5.3. Suppose $s(n) \in \text{RT}_1$ is a sequence of nonnegative terms. Then

$$s^*(1) + \cdots + ns^*(n) \in \text{RT}_1.$$

Proof. Let $S^*(n) = s^*(1) + \cdots + ns^*(n)$. Then

$$(5.1) \quad S^*(n) = \sum_{j=1}^n \left\lfloor \frac{n}{j} \right\rfloor js(j)$$

since

$$S^*(n) = \sum_{m=1}^n ms^*(m) = \sum_{jk=1}^n jk \cdot s(j)/k = \sum_{jk=1}^n js(j) = \sum_{j=1}^n \left\lfloor \frac{n}{j} \right\rfloor js(j).$$

Also,

$$(5.2) \quad S^*(n) - S^*(n-1) = ns^*(n) = \sum_{d|n} ds(d).$$

We shall show that $S^*(n) - S^*(n-1)$ is $o(S^*(n))$.

Fix $\varepsilon \in (0, 1)$ and choose

$$(5.3) \quad M > \frac{1}{\varepsilon(1-\varepsilon)}.$$

For any fixed integer r , $\frac{(n-r)s(n-r)}{ns(n)} \rightarrow 1$ as n tends to infinity. Hence we can choose $N > M^3$ such that

$$|ns(n) - (n-r)s(n-r)| < \varepsilon ns(n)$$

for $0 \leq r \leq M$ and $n \geq N/M$, and thus

$$(5.4) \quad (n-r)s(n-r) > (1-\varepsilon)ns(n)$$

for $0 \leq r \leq M$ and $n \geq N/M$.

For integers d_1, d_2 with $1 \leq d_1 < d_2 \leq M$, and for $n \geq N$, we have

$$\frac{n}{d_1} - \frac{n}{d_2} = \frac{n(d_2 - d_1)}{d_1 d_2} \geq \frac{n}{M^2} > \frac{M^3}{M^2} = M,$$

and thus $\frac{n}{d_2} < \frac{n}{d_1} - M$.

Consequently, for $n \geq N$, if $d_1 < \cdots < d_k$ are the divisors of n from the interval $[1, M]$ we see that

$$(5.5) \quad \frac{n}{M} < \frac{n}{d_k} - M < \frac{n}{d_k} < \frac{n}{d_{k-1}} - M < \cdots < \frac{n}{d_1} - M < \frac{n}{d_1}.$$

Returning to the expression for $S^*(n)$ in (5.1), now assuming that $n \geq N$, we have

$$\begin{aligned}
 S^*(n) &= \sum_{j=1}^n \left\lfloor \frac{n}{j} \right\rfloor js(j) \\
 &= \sum_{1 \leq j \leq n/M} \left\lfloor \frac{n}{j} \right\rfloor js(j) + \sum_{n/M < j \leq n} \left\lfloor \frac{n}{j} \right\rfloor js(j) \\
 &\geq \sum_{1 \leq j \leq n/M} Mjs(j) + \sum_{n/M < j \leq n} js(j) \\
 &\geq \frac{1}{\varepsilon} \sum_{1 \leq j \leq n/M} js(j) + \sum_{\substack{d|n \\ d < M}} \left(\sum_{\substack{n/d-M \leq j \\ j \leq n/d}} js(j) \right) \quad \text{by (5.3), (5.5)} \\
 &\geq \frac{1}{\varepsilon} \sum_{\substack{d|n \\ d \leq n/M}} ds(d) + \sum_{\substack{d|n \\ d < M}} M(1-\varepsilon) \frac{n}{d} \cdot s\left(\frac{n}{d}\right) \quad \text{by (5.4)} \\
 &\geq \frac{1}{\varepsilon} \sum_{\substack{d|n \\ d \leq n/M}} ds(d) + \frac{1}{\varepsilon} \sum_{\substack{d|n \\ d > n/M}} ds(d) \quad \text{by (5.3)} \\
 &= \frac{1}{\varepsilon} \sum_{d|n} ds(d) \\
 &= \frac{1}{\varepsilon} (S^*(n) - S^*(n-1)) \quad \text{by (5.2).}
 \end{aligned}$$

Thus $0 \leq S^*(n) - S^*(n-1) \leq \varepsilon S^*(n)$, and so $S^*(n) \in \text{RT}_1$. \square

6. COMBINING STAR WITH EXPONENTIATION

Theorem 6.1. Let $\mathbf{T}(x) = \exp(\mathbf{S}^*(x))$, where

- (a) $\mathbf{S}(x) \in \text{RT}_\rho$ with $0 < \rho < 1$,
- (b) the $s(n)$ are nonnegative, and
- (c) $\liminf_{n \rightarrow \infty} ns(n)\rho^n > 1$.

Then $\mathbf{T}(x) \in \text{RT}_\rho$ and $\frac{t(n-1)}{t(n)} \prec \rho$.

Furthermore, if $\mathbf{S}(\rho) = \infty$, then $s(n) = o(t(n))$.

Proof. $\mathbf{S}^*(x) \in \text{RT}_\rho$ by Lemma 5.2; so $\mathbf{T}(x) \in \text{RT}_\rho$ by Corollary 4.3. From $0 \leq s(n) \leq s^*(n)$ we have $\liminf_{n \rightarrow \infty} ns^*(n)\rho^n > 1$. Thus $t(n-1) \prec t(n)\rho$ by Lemma 4.6.

For the final assertion assume $\mathbf{S}(\rho) = \infty$. Then $\mathbf{S}^*(\rho) = \infty$; so from Lemma 4.4 we have $s^*(n) = o(t(n))$, and thus $s(n) = o(t(n))$. \square

When $\rho = 1$ we can no longer assume $\mathbf{S}^*(x) \in \text{RT}_1$ just because $\mathbf{S}(x) \in \text{RT}_1$. Instead we turn to $x\mathbf{S}'(x)/(1-x)$ to find a well-behaved sequence of coefficients.

Theorem 6.2. Let $\mathbf{T}(x) = \exp(\mathbf{S}^*(x))$ where

- (a) $\mathbf{S}(x) \in \text{RT}_1$,
- (b) the $s(n)$ are nonnegative, and
- (c) $s(n) \asymp 1/n$.

Then $\exp(\mathbf{S}^*(x)) \in \text{RT}_1$. Furthermore, $s(n) = o(t(n))$.

Proof. From

$$(6.1) \quad s^*(n) \geq s(n) \gtrsim 1/n$$

it follows that there exists a polynomial $\mathbf{p}(x)$ such that

$$(6.2) \quad \mathbf{S}^*(x) + \mathbf{p}(x) + \log(1-x)$$

has nonnegative coefficients. Then

$$(6.3) \quad \mathbf{R}(x) = \exp(\mathbf{S}^*(x) + \mathbf{p}(x))$$

has nonnegative coefficients. Since (6.2) has nonnegative coefficients it also follows that the exponential of (6.2) has nonnegative coefficients, that is,

$$(6.4) \quad [x^n]((1-x) \cdot \mathbf{R}(x)) \geq 0.$$

Differentiating (6.3), and multiplying through by x , gives

$$\begin{aligned} x\mathbf{R}'(x) &= x(\mathbf{S}^{*'}(x) + \mathbf{p}'(x)) \cdot \mathbf{R}(x) \\ &= \left(x(1-x)^{-1}(\mathbf{S}^{*'}(x) + \mathbf{p}'(x))\right) \cdot (1-x)\mathbf{R}(x). \end{aligned}$$

We will use Lemma 3.3 with

$$\begin{aligned} \mathbf{A}(x) &= x(1-x)^{-1}(\mathbf{S}^{*'}(x) + \mathbf{p}'(x)), \\ \mathbf{B}(x) &= (1-x)\mathbf{R}(x), \\ \mathbf{C}(x) &= x\mathbf{R}'(x). \end{aligned}$$

For n larger than the degree of $\mathbf{p}(x)$, we have

$$\begin{aligned} a(n) &= [x^n] \left(x(1-x)^{-1}\mathbf{S}^{*'}(x) + x(1-x)^{-1}\mathbf{p}'(x) \right) \\ (6.5) \quad &= s^*(1) + \cdots + ns^*(n) + \mathbf{p}'(1). \end{aligned}$$

From (6.1) we have $\mathbf{S}^{*'}(1) = \infty$, and thus

$$(6.6) \quad \mathbf{p}'(1) = o\left(\sum_{m < n} ms^*(m)\right).$$

From (6.5) and (6.6) it follows that

$$(6.7) \quad a(n) \sim s^*(1) + \cdots + ns^*(n).$$

So by Lemma 5.3 and Proposition 2.4, $\mathbf{A}(x) \in \text{RT}_1$. This is condition (a) of Lemma 3.3.

$\mathbf{B}(x)$ has nonnegative coefficients by (6.4), and since $\mathbf{R}(x)$ also has nonnegative coefficients,

$$0 \leq b(n) = [x^n](1-x)\mathbf{R}(x) \leq r(n) = [x^n]x\mathbf{R}'(x)/n = c(n)/n;$$

so $b(n) = o(c(n))$. This gives condition (b) of Lemma 3.3.

From (6.4) we also have

$$c(n) - c(n-1) = nr(n) - (n-1)r(n-1) \geq 0.$$

Hence condition (c) of Lemma 3.3 holds.

So, by Lemma 3.3, $\mathbf{C}(x) \in \text{RT}_1$, that is, $x\mathbf{R}'(x) \in \text{RT}_1$. Then $\mathbf{R}(x) \in \text{RT}_1$ by Proposition 2.4. So from Corollary 3.2 we have

$$\mathbf{T}(x) = \exp(-\mathbf{p}(x)) \cdot \mathbf{R}(x) \in \text{RT}_1.$$

Finally, condition (c) implies $\mathbf{S}(1) = \infty$; so

$$s(n) = o\left([x^n] \exp(\mathbf{S}(x))\right)$$

by Lemma 4.4. Now $0 \leq s(n) \leq s^*(n)$ leads to

$$[x^n] \exp(\mathbf{S}(x)) \leq [x^n] \exp(\mathbf{S}^*(x)) = [x^n] \mathbf{T}(x)$$

and thus $s(n) = o(t(n))$. □

7. COMPTON'S APPROACH TO LOGICAL LIMIT LAWS

A class \mathcal{K} of finite relational structures is said to be *adequate* if it is closed under disjoint union and components, and, up to isomorphism, it has only finitely many structures of each size. Let $p(n)$ count (up to isomorphism) the number of component structures in \mathcal{K} of size n , and let $a(n)$ count the total number of structures in \mathcal{K} of size n . The combinatorial identity connecting the two counting functions $a(n)$ and $p(n)$ is

$$(7.1) \quad \mathbf{A}(x) = \exp(\mathbf{P}^*(x)),$$

where

$$\begin{aligned} \mathbf{A}(x) &= \sum a(n)x^n, & \mathbf{P}^*(x) &= \sum p^*(n)x^n, \\ p^*(0) &= 0, & p^*(n) &= \sum_{j+k=n} p(j)/k \quad \text{for } n \geq 1. \end{aligned}$$

The connection between adequate classes and (7.1) is very tight, for if $p(n)$ is any nonnegative integer-valued function with $p(0) = 0$, then there is an adequate class \mathcal{K} with $p(n)$ the count function for the components of \mathcal{K} , and the function $a(n)$ satisfying (7.1) is the total count function for \mathcal{K} .

Compton proved two main theorems for the purpose of finding classes \mathcal{K} with logical limit laws. We assume that \mathcal{K} is an adequate class of relational structures with the counting functions $a(n)$ and $p(n)$ as described above. Furthermore, we assume $a(n) \succ 0$.² The striking feature of Compton's theorems is that he is able to prove logical limit laws just from knowing information about the counting function $a(n)$. Note that if $a(n) \in \text{RT}_\rho$, then $0 \leq \rho \leq 1$ since the $a(n)$ are integers. We only consider $0 < \rho \leq 1$ since Compton's method does not work for the case $\rho = 0$. (The case $\rho = 0$ requires more knowledge about \mathcal{K} than just the count functions to determine if there is a logical limit law.)

Theorem 7.1 (Compton, 1987/1989). *If $a(n) \in \text{RT}_1$, then \mathcal{K} has a monadic second-order zero-one law.*

Theorem 7.2 (Compton, 1989). *If $a(n) \in \text{RT}_\rho$, where $0 < \rho < 1$, and if there exist K and C such that*

$$(7.2) \quad \frac{a(n-k)}{a(n)} \leq C\rho^k \quad \text{for } K \leq k \leq n,$$

then \mathcal{K} has a monadic second-order limit law.

²This is the same as requiring that the gcd of the sizes of the components be 1.

8. A USEFUL COROLLARY

For applications of Theorem 7.2 we use the following.

Corollary 8.1. *Suppose $a(n) \in \text{RT}_\rho$ and $\frac{a(n-1)}{a(n)} \preccurlyeq \rho$. Then \mathcal{K} has a monadic second-order limit law.*

Proof. Choose $K \geq 1$ such that

$$n \geq K \implies \begin{cases} a(n) > 0, & \text{and} \\ \frac{a(n-1)}{a(n)} \leq \rho. \end{cases}$$

Now suppose $K \leq k \leq n$.

Case 1: If $n - k \geq K$, then

$$\frac{a(n-k)}{a(n)} = \frac{a(n-k)}{a(n-k+1)} \cdots \frac{a(n-1)}{a(n)} \leq \rho^k.$$

Case 2: If $n - k < K$, then

$$\begin{aligned} \frac{a(n-k)}{a(n)} &= \frac{a(n-k)}{a(K)} \frac{a(K)}{a(n)} \\ &\leq \frac{a(n-k)}{a(K)} \rho^{n-K} && \text{by Case 1} \\ &= \left(\frac{a(n-k)}{a(K)} \rho^{n-K-k} \right) \rho^k \\ &\leq \left(\frac{1}{a(K)} \cdot \max(a(0), \dots, a(K-1)) \cdot \rho^{-K} \right) \rho^k. \end{aligned}$$

Thus Theorem 7.2 applies. □

9. APPLICATIONS TO LOGICAL LIMIT LAWS

Now we use our results to improve the scope of the conditions on $p(n)$ that lead to classes of structures covered by Compton's theorems.

Theorem 9.1. *If $p(n) \in \text{RT}_1$, then \mathcal{K} has a monadic second-order zero-one law.*

Proof. This follows from Theorems 6.2 and 7.1 after noting that $p(n)$ is always a nonnegative integer. □

This theorem yields many classes of structures that, before, were not known to have a 0–1 law. For example, if

$$p(n) \sim an^b e^{cn^d},$$

with $d < 1$ and $a, c > 0$, then one has a monadic second-order 0–1 law. Previously the known comprehensive general collection of count functions $p(n)$ that implied such a law was the family of polynomially bounded $p(n)$, that is, $p(n) = O(n^c)$ for some c (see Bell [1]). Now we have entire families of superpolynomial count functions $p(n)$ that guarantee such laws.

Example 9.2. Let \mathcal{K} be the class of finite directed graphs $\mathbf{G} = (G, R)$ satisfying the first-order conditions

$$\begin{aligned} &\forall x (\exists y (yRx) \leftrightarrow xRx), \\ &\forall x \forall y ((xRx) \& (xRy) \rightarrow (yRx)), \\ &\forall x \forall y \forall z ((xRx) \& (xRy) \& (yRz) \rightarrow (xRz)), \\ &\forall x (\neg(xRx) \rightarrow \exists! y (xRy)). \end{aligned}$$

Such a digraph consists of an equivalence relation plus a possibly empty collection of vertices that each have but one edge attached to them, and that edge is outgoing to an element in the equivalence relation. The vertices comprising the equivalence relation are precisely those in the range of the binary relation R . The indecomposable members of \mathcal{K} are those with a single equivalence class. It is easy to see that $p(n)$ is the number of partitions of the integer n , which we know from the famous results of Hardy and Ramanujan to satisfy

$$p(n) \sim \frac{1}{4\sqrt{3}} n^{-1} e^{\sqrt{2/3}\pi n^{1/2}}.$$

By Theorem 9.1 we see that \mathcal{K} has a monadic second-order 0–1 law.

Theorem 9.3. Suppose $0 < \rho < 1$. If

- (a) $p(n) \in \text{RT}_\rho$ and
- (b) $\liminf_{n \rightarrow \infty} np(n)\rho^n > 1$,

then \mathcal{K} has a monadic second-order limit law.

Proof. From Theorem 6.1 we know $a(n) \in \text{RT}_\rho$ as well as $\frac{a(n-1)}{a(n)} \prec \rho$; so Corollary 8.1 applies. \square

Prior to Theorem 9.3 all applications of Compton's Theorem 7.2 were based on the asymptotics of Knopfmacher, Knopfmacher, and Warlimont [11] that say if

$$(9.1) \quad p(n) = C_1 \beta^n + O(\gamma^n), \quad 0 < \gamma < \beta, \quad C_1 > 0,$$

then

$$a(n) \sim C_2 \beta^n e^{2\sqrt{an}} / n^{3/4}, \quad \text{for some } C_2 > 0.$$

These results were proved by the well-known use of Cauchy's integral theorem for asymptotics. In this case,

$$a(n) = \frac{1}{2\pi i} \int_C \exp(\mathbf{P}^*(z)) \cdot \frac{dz}{z^{n+1}}.$$

From the hypothesis (9.1) we immediately have $\rho = 1/\beta$ and $p(n) \in \text{RT}_\rho$ as well as $\liminf_{n \rightarrow \infty} np(n)\rho^n > 1$. Thus Theorem 9.3 covers condition (9.1). But we can say much more. In the following corollary, we focus on generalizing the $C\beta^n$ asymptotics and drop any reference to an error term.

Corollary 9.4. For $\mu < 1 < \beta$ and $C > 0$, or $\mu = 1 < \beta$ and $C > 1$, the condition

$$p(n) \sim C\beta^n / n^\mu$$

guarantees that \mathcal{K} has a monadic second-order limit law.

Proof. From Proposition 2.4 we see that $p(n) \in \text{RT}_\rho$, where $\rho = 1/\beta$, and the condition $\liminf_{n \rightarrow \infty} np(n)\rho^n > 1$ is readily verified. Thus Theorem 9.3 applies. \square

Example 9.5. For a fixed positive integer k , let \mathcal{K} be the class of finite k -colored linear forests. (The components of linear forests are chains. If $k = 4$, then one can think of a member of \mathcal{K} as a collection of DNA fragments.) Then

$$p(n) \sim k^n;$$

so by Corollary 9.4 we know that \mathcal{K} has a monadic second-order limit law.

Example 9.6. For h a fixed positive integer, let \mathcal{K} be the class of finite forests of planted plane trees of height at most h , that is, the class of finite graphs whose components are planted plane trees of height at most h .

The basic study of the generating function $\mathbf{P}(x) := \sum p(n)x^n$ for finite planted plane trees can be found in the 1972 paper [9] of de Bruijn, Knuth and Rice. From their work we know that

$$\mathbf{P}(x) = \frac{f(x)}{g(x)}$$

where $f(x)$ and $g(x)$ are the polynomials given by

$$\begin{aligned} f(x) &= 2x \cdot \frac{(1 + \sqrt{1-4x})^h - (1 - \sqrt{1-4x})^h}{\sqrt{1-4x}}, \\ g(x) &= \frac{(1 + \sqrt{1-4x})^{h+1} - (1 - \sqrt{1-4x})^{h+1}}{\sqrt{1-4x}}. \end{aligned}$$

They also note that the degree of $g(x)$ is $d = \lfloor h/2 \rfloor$, and that $g(x)$ has d distinct positive roots r_j given by

$$r_j = \frac{1}{4} \sec^2 \left(\frac{j\pi}{h+1} \right) \quad \text{for } 1 \leq j \leq d.$$

Of course we need $h \geq 2$ in order for $g(x)$ to have any roots. Clearly $0 < r_1 < \dots < r_d$, and for $h \geq 3$ the smallest root r_1 lies in the interval $(1/4, 1/2]$.

Given that $h \geq 3$, for some $c \neq 0$ we have

$$g(x) = c(x - r_1) \cdots (x - r_d).$$

So by the method of partial fractions,

$$\frac{1}{g(x)} = \sum_{j=1}^d \frac{c_j}{x - r_j} \quad \text{where } c_j = c^{-1} \prod_{i \neq j} (r_j - r_i)^{-1}.$$

From this we easily have

$$\begin{aligned} p(n) &= [x^n] \frac{f(x)}{g(x)} \\ &= - \sum_{j=1}^d c_j f(r_j) / r_j^{n+1} \quad \text{for } n \geq \left\lfloor \frac{h-1}{2} \right\rfloor (= \deg(f)) \\ &\sim -c_1 \frac{f(r_1)}{r_1} \left(\frac{1}{r_1} \right)^n, \end{aligned}$$

and thus for $h \geq 3$ the class \mathcal{K} has a monadic second-order limit law by Corollary 9.4. In the cases $h = 1, 2$ the values $p(n)$ are uniformly bounded, and thus one has a monadic second-order 0-1 law.

Remark 9.7. If we change the condition $\mu < 1$ in Corollary 9.4 to $\mu > 1$, then we can find examples of classes \mathcal{K} with such a count function $p(n)$ that fail to have a first-order limit law.

Remark 9.8. We can construct, for any given ρ with $0 < \rho \leq 1$, an infinite sequence \mathcal{K}_m of classes of finite relational structures with monadic second-order limit laws such that the count functions $p_m(n)$ and $a_m(n)$ for \mathcal{K}_m are in RT_ρ and growing infinitely faster at each successive step; that is, we have

$$p_m(n) = o(a_m(n)) \quad \text{and} \quad a_m(n) = o(p_{m+1}(n)).$$

To construct the sequence we start with

$$p_0(n) = \lfloor 1/\rho^n \rfloor, \quad \text{for } n \geq 1.$$

Then

$$a_0(n) = \lfloor x^n \rfloor \exp(\mathbf{P}_0^*(x))$$

gives $p_0(n) = o(a_0(n))$ by Theorem 6.1 or 6.2. Now observe that by modifying $a_0(n)$, by simply setting $a_0(0)$ to 0, we have a sequence that can be used as a $p(n)$, satisfying the premises of Theorem 6.1 or 6.2. Inductively define p_{k+1} and a_{k+1} by: $p_{k+1}(0) = 0$ and

$$p_{k+1}(n) = \lfloor x^n \rfloor \exp(\mathbf{A}_k^*(x)), \quad \text{for } n \geq 1,$$

$$a_{k+1}(n) = \lfloor x^n \rfloor \exp(\mathbf{P}_{k+1}^*(x)), \quad \text{for } n \geq 0,$$

where $\mathbf{P}_k(x) = \sum_n p_k(n)x^n$ and $\mathbf{A}_k(x) = \sum_n a_k(n)x^n$. Now take classes \mathcal{K}_m with counting functions $a_m(n)$ and $p_m(n)$.

In the case that $\rho = 1$ these classes have, for $m \geq 1$, both counting functions with superpolynomial growth, growing far faster than the classes with polynomially bounded $p(n)$ that were, with minor exceptions, the only ones that were previously known to have a monadic second-order 0–1 law. Likewise, with $0 < \rho < 1$, the classes \mathcal{K}_m , for $m \geq 1$, grow far faster than any examples that we knew before with a monadic second-order limit law.

REFERENCES

- [1] Jason P. Bell, *Sufficient conditions for zero-one laws*. Trans. Amer. Math. Soc. **354** (2002), no. 2, 613–630. MR **2002j**:60057
- [2] Jason P. Bell, Edward A. Bender, Peter J. Cameron, and L. Bruce Richmond, *Asymptotics for the probability of connectedness and the distribution of number of components*. Electron. J. Combin. **7** (2000), R33. MR **2001d**:05009
- [3] Edward A. Bender, *Asymptotic methods in enumeration*. SIAM Rev. **16** (1974), 485–515. MR **51**:12545
- [4] Stanley N. Burris, *Number Theoretic Density and Logical Limit Laws*. Mathematical Surveys and Monographs, Vol. **86**, Amer. Math. Soc., Providence, RI, 2001. MR **2002c**:03060
- [5] Stanley Burris, *Spectrally determined first-order limit laws*. Logic and Random Structures (New Brunswick, NJ, 1995), 33–52, ed. by Ravi Boppana and James Lynch. DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **33**, Amer. Math. Soc., Providence, RI, 1997. MR **98j**:03047
- [6] Stanley Burris and András Sárközy, *Fine spectra and limit laws I. First-order laws*. Canad. J. Math. **49** (1997), 468–498. MR **98k**:03076a
- [7] Kevin J. Compton, *A logical approach to asymptotic combinatorics. I. First order properties*. Adv. in Math. **65** (1987), 65–96. MR **88k**:03065
- [8] Kevin J. Compton, *A logical approach to asymptotic combinatorics. II. Monadic second-order properties*. J. Combin. Theory, Ser. A **50** (1989), 110–131. MR **90c**:03024
- [9] N. G. de Bruijn, D. E. Knuth, and S. O. Rice, *The average height of planted plane trees*. Graph Theory and Computing, Academic Press, New York, 1972, pp. 15–22. MR **58**:21737

- [10] R. Durrett, B. Granovsky, and S. Gueron, *The equilibrium behavior of reversible coagulation-fragmentation processes*. J. Theoret. Probab. **12** (1999), 447–474. MR **2000g**:82013
- [11] Arnold Knopfmacher, John Knopfmacher, and Richard Warlimont, “*Factorisatio numerorum*” in *arithmetical semigroups*. Acta Arith. **61** (1992), 327–336. MR **93f**:11069

MATHEMATICS DEPARTMENT, UNIVERSITY OF MICHIGAN, EAST HALL, 525 EAST UNIVERSITY,
ANN ARBOR, MICHIGAN 48109-1109

E-mail address: belljp@umich.edu

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO N2L
3G1 CANADA

E-mail address: snburris@thoralf.uwaterloo.ca

URL: www.thoralf.uwaterloo.ca

COMBINATORICS OF ROOTED TREES AND HOPF ALGEBRAS

MICHAEL E. HOFFMAN

ABSTRACT. We begin by considering the graded vector space with a basis consisting of rooted trees, with grading given by the count of non-root vertices. We define two linear operators on this vector space, the growth and pruning operators, which respectively raise and lower grading; their commutator is the operator that multiplies a rooted tree by its number of vertices, and each operator naturally associates a multiplicity to each pair of rooted trees. By using symmetry groups of trees we define an inner product with respect to which the growth and pruning operators are adjoint, and obtain several results about the associated multiplicities.

Now the symmetric algebra on the vector space of rooted trees (after a degree shift) can be endowed with a coproduct to make a Hopf algebra; this was defined by Kreimer in connection with renormalization. We extend the growth and pruning operators, as well as the inner product mentioned above, to Kreimer's Hopf algebra. On the other hand, the vector space of rooted trees itself can be given a noncommutative multiplication: with an appropriate coproduct, this leads to the Hopf algebra of Grossman and Larson. We show that the inner product on rooted trees leads to an isomorphism of the Grossman-Larson Hopf algebra with the graded dual of Kreimer's Hopf algebra, correcting an earlier result of Panaite.

1. INTRODUCTION

In recent work on renormalization of quantum field theory, D. Kreimer and his collaborators [1], [3], [13], [14], [15], [16] introduce a Hopf algebra (here denoted \mathcal{H}_K) whose generators are rooted trees. Various other Hopf algebras based on rooted trees have appeared in the literature, in particular that of R. Grossman and R. G. Larson [10], which F. Panaite [18] connected to \mathcal{H}_K . The proof of the principal result of [18] actually contains an error due to the confusion of two kinds of multiplicities associated to triples of rooted trees. In this paper we show how to correct Panaite's result, while clarifying the combinatorial significance of these multiplicities.

Kreimer's Hopf algebra \mathcal{H}_K admits a derivation called the growth operator, which is important in describing the relation of this algebra to another Hopf algebra studied earlier by A. Connes and H. Moscovici [4]. We introduce a complementary derivation called the pruning operator. In fact, we find it easiest to start (in §2) in

Received by the editors June 25, 2002 and, in revised form, February 24, 2003.

2000 *Mathematics Subject Classification.* Primary 05C05, 16W30; Secondary 81T15.

Key words and phrases. Rooted tree, Hopf algebra, differential poset.

The author was partially supported by a grant from the Naval Academy Research Council.

Some of the results of this paper were presented to an AMS Special Session on Combinatorial Hopf Algebras on May 4, 2002.

the vector space $k\{\mathcal{T}\}$ of rooted trees rather than in \mathcal{H}_K . There we have growth and pruning operators (denoted by \mathfrak{N} and \mathfrak{P} respectively), and for each pair of rooted trees t, t' with $|t| \leq |t'|$ (where $|t|$ is the number of vertices of t) there are natural multiplicities $n(t; t')$ and $m(t; t')$ associated with \mathfrak{N} and \mathfrak{P} respectively. A comparison of these multiplicities using symmetry groups of rooted trees leads to the definition of an inner product with respect to which \mathfrak{N} and \mathfrak{P} are adjoint. The operators \mathfrak{N} and \mathfrak{P} are very similar to the adjoint operators that appear in Stanley’s theory of differential posets [21], [22], and can be described in terms of Fomin’s somewhat more general theory of dual graded graphs [6], [7], [8]. The techniques of Stanley and Fomin are used to obtain various results about n and m .

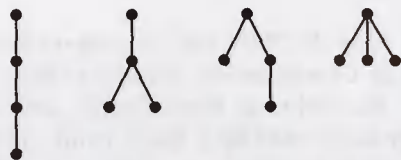
In §3 we extend the growth and pruning operators (as well as the inner product) to the Hopf algebra \mathcal{H}_K . In this setting the growth and pruning operators are not quite adjoint, but the deviation from adjointness is easily described (Proposition 3.3). We also describe the duals of the growth and pruning operators. In §4 we extend the multiplicities n and m to multiplicities $n(t_1, t_2; t_3)$ and $m(t_1, t_2; t_3)$ associated to triples of rooted trees t_1, t_2, t_3 with $|t_1| + |t_2| = |t_3|$. We then give an explicit isomorphism of the Grossman-Larson Hopf algebra onto the graded dual of \mathcal{H}_K (Proposition 4.4), providing a corrected version of Panaite’s result. We also show how the isomorphism gives another description of the duals of the growth and pruning operators.

In addition to the references cited above, two recent articles treat aspects of \mathcal{H}_K not covered here: see [2] for connections between Kreimer’s Hopf algebra and earlier work on Runge-Kutta methods, and [5] for an analysis of the primitives of \mathcal{H}_K . The author thanks the referee for bringing them to his attention.

2. THE VECTOR SPACE OF ROOTED TREES

A rooted tree is a partially ordered set (whose elements are called vertices) with a unique greatest element (the root vertex), such that, for any vertex v , the vertices exceeding v in the partial order form a chain. If v exceeds w in the partial order, we call w a descendant of v and v an ancestor of w . If v covers w in the partial order (i.e., v is an ancestor of w and there are no vertices between v and w in the order), we call w a child of v and v the parent of w .

We can visualize a rooted tree as a directed graph by putting an edge from each vertex to each of its children: the root is the only vertex with no incoming edge. We call a vertex terminal if it has no outgoing edges (i.e., no children). The condition that the set of ancestors of any vertex forms a chain insures that this graph has no cycles. For a finite rooted tree t , we denote by $|t|$ the number of vertices of t : let \mathcal{T} be the set of finite rooted trees, and $\mathcal{T}_n = \{t \in \mathcal{T} : |t| = n + 1\}$. For example, $\mathcal{T}_0 = \{\bullet\}$, where \bullet is the tree consisting of only the root vertex, and below are the four elements of \mathcal{T}_3 , with the root placed at the top:



We can define a partial order \preceq on the set \mathcal{T} itself by setting $t \preceq t'$ if t can be obtained from t' by removing some non-root vertices and edges; of course $t \preceq t'$ implies $|t| \leq |t'|$. Evidently t' covers t for this order exactly in the case when t can

be obtained by removing from t' a single terminal vertex and the edge into it. In this case we write $t \triangleleft t'$.

If $t \triangleleft t'$, we can get from t' to t by removing a (terminal) edge, and from t to t' by adding an edge (and accompanying terminal vertex). This leads to the definitions of two numbers associated with the pair (t, t') :

$$n(t; t') = |\{\text{vertices of } t \text{ to which a new edge can be added to get } t'\}|$$

and

$$m(t; t') = |\{\text{edges of } t' \text{ which when removed leave } t\}|.$$

That these two numbers are not always equal can be seen from the example of



where $n(t; t') = 1$ and $m(t; t') = 2$.

The relation between the numbers $n(t; t')$ and $m(t; t')$ can be clarified by introducing symmetry groups of trees. For a rooted tree t , let $V(t)$ be its set of vertices: then for each $v \in V(t)$, there is a rooted tree t_v consisting of v and its descendants with the order inherited from t . We call this the subtree of t with v as root. For $v \in V(t)$, let $SG(t, v)$ be the group of permutations of identical branches out of v , i.e., if $\{v_1, v_2, \dots, v_k\}$ are the children of v , then $SG(t, v)$ is the group generated by the permutations that exchange t_{v_i} with t_{v_j} when they are isomorphic rooted trees. The symmetry group of t is the direct product

$$SG(t) = \prod_{v \in V(t)} SG(t, v).$$

For a given $v \in V(t)$, let $\text{Fix}(v, t) \leq SG(t)$ be the subgroup of $SG(t)$ that fixes v ; note that $\text{Fix}(w, t) \leq \text{Fix}(v, t)$ whenever w is a descendant of v .

Now suppose $t \triangleleft t'$, and let $v \in V(t)$ be such that, when a new edge and terminal vertex w are added to t at v , the result is t' . If $\text{Orb}(v, t)$ is the orbit of v under $SG(t)$, then evidently

$$n(t; t') = |\text{Orb}(v, t)| = |SG(t) / \text{Fix}(v, t)| = \frac{|SG(t)|}{|\text{Fix}(v, t)|}.$$

On the other hand, if $\text{Orb}(w, t')$ is the orbit of $w \in V(t')$ under $SG(t')$, then

$$m(t; t') = |\text{Orb}(w, t')| = |SG(t') / \text{Fix}(w, t')| = \frac{|SG(t')|}{|\text{Fix}(w, t')|}.$$

But there is an evident identification of $\text{Fix}(v, t)$ with $\text{Fix}(w, t')$; so we have the following result.

Proposition 2.1. *If $t \triangleleft t'$, then $|SG(t)|m(t; t') = n(t; t')|SG(t')|$.*

Let

$$k\{\mathcal{T}\} = \bigoplus_{n \geq 0} k\{\mathcal{T}_n\}$$

be the graded vector space (over a field k of characteristic 0) with basis consisting of rooted trees: we put the rooted tree t in grade $|t| - 1$. We define two linear operators

on $k\{\mathcal{T}\}$ as follows. For $n \geq 0$, the growth operator $\mathfrak{N} : k\{\mathcal{T}_n\} \rightarrow k\{\mathcal{T}_{n+1}\}$ is defined by

$$(1) \qquad \mathfrak{N}(t) = \sum_{t' \triangleleft t'} n(t; t') t',$$

and for $n \geq 1$ the pruning operator $\mathfrak{P} : k\{\mathcal{T}_n\} \rightarrow k\{\mathcal{T}_{n-1}\}$ is given by

$$(2) \qquad \mathfrak{P}(t) = \sum_{t' \triangleleft t} m(t'; t) t';$$

we set $\mathfrak{P}(\bullet) = 0$. Then \mathfrak{P} and \mathfrak{N} satisfy the following commutation relation.

Proposition 2.2. *As operators on $k\{\mathcal{T}\}$, $[\mathfrak{P}, \mathfrak{N}] = \mathfrak{D}$, where \mathfrak{D} is the operator given by $\mathfrak{D}(t) = |t|t$.*

Proof. It suffices to show $\mathfrak{P}\mathfrak{N}(t) - \mathfrak{N}\mathfrak{P}(t) = |t|t$ for any rooted tree t . Let $V(t) = \{v_1, \dots, v_n, v_{n+1}, \dots, v_{|t|}\}$ be the vertices of t , with v_i terminal for $1 \leq i \leq n$. Then

$$\mathfrak{N}(t) = \sum_{i=1}^{|t|} t_i \quad \text{and} \quad \mathfrak{P}(t) = \sum_{i=1}^n t^{(i)},$$

where t_i is the tree obtained from t by adding a new edge and terminal vertex to t at v_i , and $t^{(i)}$ comes from t by removing the edge that ends in v_i . Then $\mathfrak{P}\mathfrak{N}(t)$ is

$$\sum_{i=1}^{|t|} \mathfrak{P}(t_i) = \sum_{i=1}^{|t|} \left(t + \sum_{1 \leq j \leq n, j \neq i} (t_i)^{(j)} \right) = |t|t + \sum_{i=1}^{|t|} \sum_{1 \leq j \leq n, j \neq i} (t_i)^{(j)},$$

while $\mathfrak{N}\mathfrak{P}(t)$ is

$$\sum_{j=1}^n \mathfrak{N}(t^{(j)}) = \sum_{i=1}^{|t|} \sum_{1 \leq j \leq n, j \neq i} (t^{(j)})_i.$$

Since $(t_i)^{(j)} = (t^{(j)})_i$ for $i \neq j$, the conclusion follows. □

Now we can endow $k\{\mathcal{T}\}$ with an inner product by setting

$$(t, t') = |SG(t)|\delta_{t, t'}$$

for any rooted trees t, t' .

Proposition 2.3. *The operators \mathfrak{N} and \mathfrak{P} are adjoint with respect to the inner product (\cdot, \cdot) .*

Proof. It suffices to show

$$(\mathfrak{N}(t), t') = (t, \mathfrak{P}(t'))$$

when $t \triangleleft t'$ (otherwise both sides are zero). In this case, we have

$$(\mathfrak{N}(t), t') = n(t; t')(t', t') = n(t; t')|SG(t')|$$

from equation (1) and

$$(t, \mathfrak{P}(t')) = m(t; t')(t, t) = m(t; t')|SG(t)|$$

from equation (2); but then the result follows by Proposition 2.1. □

Putting the last two results together gives the following.

Proposition 2.4. *For rooted trees t_1 and t_2 ,*

$$(\mathfrak{N}(t_1), \mathfrak{N}(t_2)) - (\mathfrak{P}(t_1), \mathfrak{P}(t_2)) = \begin{cases} 0, & \text{if } t_1 \neq t_2, \\ |t||SG(t)|, & \text{if } t_1 = t_2 = t. \end{cases}$$

Proof. This follows from the calculation

$$(\mathfrak{N}(t_1), \mathfrak{N}(t_2)) - (\mathfrak{P}(t_1), \mathfrak{P}(t_2)) = (t_1, (\mathfrak{P}\mathfrak{N} - \mathfrak{N}\mathfrak{P})(t_2)) = (t_1, \mathfrak{D}(t_2)) = |t_2|(t_1, t_2).$$

□

Remark. The second alternative of this result can be written

$$\sum_{t \triangleleft t'} n(t; t')^2 |SG(t')| - \sum_{t'' \triangleleft t} m(t''; t)^2 |SG(t'')| = |t||SG(t)|,$$

or, dividing by $|SG(t)|$,

$$\sum_{t \triangleleft t'} n(t; t')m(t; t') - \sum_{t'' \triangleleft t} n(t''; t)m(t''; t) = |t|$$

for any rooted tree t .

We can extend the definitions of $m(t; t')$ and $n(t; t')$ to any pair of rooted trees t, t' with $|t'| - |t| = k \geq 0$ by setting

$$(3) \quad \mathfrak{N}^k(t) = \sum_{|t'|=|t|+k} n(t; t')t'$$

and

$$(4) \quad \mathfrak{P}^k(t') = \sum_{|t|=|t'|+k} m(t; t')t.$$

With these definitions, we have the following result.

Proposition 2.5. *Let t, t' be rooted trees with $|t| \leq |t'|$. Then*

1. $n(t; t')|SG(t')| = |SG(t)|m(t; t')$.
2. *If $|t| \leq k \leq |t'|$, then*

$$n(t; t') = \sum_{|t''|=k} n(t; t'')n(t''; t'),$$

and similarly for n replaced by m .

3. $n(t; t')$ and $m(t; t')$ are nonzero if and only if $t \preceq t'$.

Proof. The first part follows immediately from equations (3) and (4):

$$n(t; t')|SG(t')| = (t', \mathfrak{N}^{|t'| - |t|}(t)) = (\mathfrak{P}^{|t'| - |t|}(t'), t) = m(t; t')|SG(t)|.$$

For the second part, we have for $|t| \leq k \leq |t'|$,

$$\begin{aligned} n(t; t') &= \frac{(\mathfrak{N}^{|t'|-|t|}(t), t')}{|SG(t')|} \\ &= \frac{(\mathfrak{N}^{k-|t|}(t), \mathfrak{P}^{|t'|-k}(t'))}{|SG(t')|} \\ &= \sum_{|t''|=k} \frac{(\mathfrak{N}^{k-|t|}(t), m(t''; t')t'')}{|SG(t')|} \\ &= \sum_{|t''|=k} \frac{(\mathfrak{N}^{k-|t|}(t), t'')}{|SG(t'')|} \frac{|SG(t'')|}{|SG(t')|} m(t''; t') \\ &= \sum_{|t''|=k} n(t; t'') n(t''; t'). \end{aligned}$$

(For the corresponding equation with m replacing n , reverse the roles of \mathfrak{N} and \mathfrak{P} .) Finally, the third part is evident for $|t'| - |t| = 1$ and can be proved by induction on $|t'| - |t|$ using the second part. \square

Remark. The second and third parts say that \mathcal{T} is a weighted-relation poset, in the terminology of [11], for either of the weights $n(t; t')$ or $m(t; t')$. In fact, \mathcal{T} with weights $n(t; t')$ is discussed as Example 7 in [11]. In the terminology of [7], \mathcal{T} with multiplicities $n(t; t')$ and \mathcal{T} with multiplicities $m(t; t')$ are a pair of graded graphs that are \mathbf{r} -dual for the sequence $\mathbf{r} = (0, 1, 2, \dots)$.

If $t \preceq t'$, we can think of $n(t; t')$ as counting the ways of building up t' from t by adding new edges and terminal vertices, and $m(t; t')$ as counting ways of getting from t' to t by removing terminal edges. In particular, since $\bullet \preceq t$ for every rooted tree t , we can think of $n(\bullet; t)$ as the number of ways to build up t , and $m(\bullet; t)$ as the number of ways to tear it down. A more precise formulation can be given using the idea of labellings of trees: a labelling of a rooted tree t is a bijection $f : V(t) \rightarrow \{0, 1, \dots, |t|\}$ such that $f(v) > f(w)$ whenever v is a descendant of w (necessarily f sends the root vertex to 0). We call labellings f and g equivalent if $f\phi = g$ for some $\phi \in SG(t)$.

Proposition 2.6. *Let t be a rooted tree. Then t has $m(\bullet; t)$ labellings and $n(\bullet; t)$ labellings mod equivalence.*

Proof. First note that $|SG(t)|n(\bullet; t) = m(\bullet; t)$ by the first part of Proposition 2.5 since \bullet has trivial symmetry group. It follows from the discussion of [11, Ex. 7] that $n(\bullet; t)$ counts labellings mod equivalence, and the statement about $m(\bullet; t)$ follows since each equivalence class of labellings has $|SG(t)|$ elements. \square

Remark. The “Connes-Moscovici weight” [1], [15] or “tree multiplicity” [2] of t is $n(\bullet; t)$. Cf. [20, Sect. 22] and [12, Ex. 5.1.4-20], where a hook-length formula for the number of labellings of t is given: this is $m(\bullet; t)$.

In [22] Stanley defined the notion of a sequentially differential poset (generalizing his definition of a differential poset in [21]). A sequentially differential poset P is a locally finite graded poset with a single element $\hat{0}$ in grade 0, so that the linear operators

$$U(p) = \sum_{p \triangleleft p'} p' \quad \text{and} \quad D(p) = \sum_{p' \triangleleft p} p'$$

on $k\{P\}$ satisfy the identity $(DU - UD)(p) = r_j p$ for any $p \in P$ of rank j : here r_0, r_1, \dots are nonnegative integers. The results of [22] can be applied to \mathcal{T} (with $r_j = j + 1$), provided we replace U and D with \mathfrak{N} and \mathfrak{P} respectively, and suitably reinterpret the statements of theorems to incorporate multiplicities. For example, for $x \in P$ Stanley writes $e(x)$ for the number of saturated chains from $\hat{0}$ to x , but in the proofs $e(x)$ really appears as the inner product of x with $U^k \hat{0}$, where k is the rank of x : so for a rooted tree $x \in \mathcal{T}_k$ we replace $e(x)$ by

$$(\mathfrak{N}^k \bullet, x) = n(\bullet; x)(x, x) = n(\bullet; x)|SG(x)| = m(\bullet; x).$$

Let $w = w_1 w_2 \cdots w_r$ be a word in \mathfrak{N} and \mathfrak{P} , and let $x \in \mathcal{T}_k$. Clearly $(w \bullet, x) = 0$ unless (a) for each $1 \leq i \leq r$, the number of \mathfrak{P} 's in $w_i w_{i+1} \cdots w_r$ does not exceed the number of \mathfrak{N} 's; and (b) the number of \mathfrak{N} 's minus the number of \mathfrak{P} 's in w is k . In this case we call w a valid x -word, and we have the following result.

Proposition 2.7. *Let $x \in \mathcal{T}_k$, $w = w_1 \cdots w_r$ a valid x -word. Let $S = \{i : w_i = \mathfrak{P}\}$. For each $i \in S$, let $a_i = |\{j : j \geq i, w_j = \mathfrak{P}\}|$, $b_i = |\{j : j > i, w_j = \mathfrak{N}\}|$, and $c_i = b_i - a_i$. Then*

$$(w \bullet, x) = m(\bullet; x) \prod_{i \in S} \binom{c_i + 2}{2}.$$

Proof. Replace U, D by $\mathfrak{N}, \mathfrak{P}$ in Theorem 2.3 of [22]. □

This result has the following corollary (cf. Theorem 1.5.2 of [7]).

Proposition 2.8. *For any rooted tree $x \in \mathcal{T}_k$ and nonnegative integer a ,*

$$\sum_{|t|=k+a+1} m(x; t) n(\bullet; t) = n(\bullet; x) \prod_{i=2}^{a+1} \binom{k+i}{2}.$$

Proof. In Proposition 2.7 set $w = \mathfrak{P}^a \mathfrak{N}^{a+k}$ to get

$$(\mathfrak{P}^a \mathfrak{N}^{a+k} \bullet, x) = m(\bullet; x) \prod_{i=2}^{a+1} \binom{k+i}{2}.$$

Now the left-hand side can be expanded as

$$\begin{aligned} (\mathfrak{N}^{a+k} \bullet, \mathfrak{N}^a(x)) &= \sum_{|t|=k+a+1} n(x; t) (\mathfrak{N}^{a+k} \bullet, t) = \sum_{|t|=k+a+1} n(x; t) m(\bullet; t) \\ &= \sum_{|t|=k+a+1} m(x; t) |SG(x)| n(\bullet; t), \end{aligned}$$

where we have the first part of Proposition 2.5 in the last step. Hence

$$\sum_{|t|=k+a+1} m(x; t) |SG(x)| n(\bullet; t) = m(\bullet; x) \prod_{i=2}^{a+1} \binom{k+i}{2},$$

and dividing by $|SG(x)|$ gives the conclusion. □

Remark. In the case $x = \bullet$, this result becomes

$$\sum_{|t|=a+1} m(\bullet; t) n(\bullet; t) = \sum_{|t|=a+1} n(\bullet; t)^2 |SG(t)| = \prod_{i=2}^{a+1} \binom{i}{2}.$$

Cf. [7, Cor. 1.5.4].

In [11, Ex. 7] it is shown that $\sum_{|t|=k+1} n(\bullet; t) = k!$. Further sum formulas involving $n(\bullet; t)$ appear in [15, Sect. 5] and [2, Sect. 5]. A result of [22] gives a formula for $\sum_{|t|=k+1} m(\bullet; t)$. To state it we will need some definitions. Let $\text{Inv}(k)$ be the set of involutions in the group Σ_k of permutations of $\{1, 2, \dots, k\}$. For $\sigma \in \Sigma_k$, call i a weak excedance of σ if $\sigma(i) \geq i$; let $\text{Wex}(\sigma)$ be the set of weak excedances of σ . For $\sigma \in \Sigma_k$ and $i \in \{1, \dots, k\}$, let $\eta(\sigma, i)$ be the number of integers j such that $j < i$ and $\sigma(j) < \sigma(i)$. Then we have the following result.

Proposition 2.9. *With the definitions above,*

$$\sum_{|t|=k+1} m(\bullet; t) = \sum_{\sigma \in \text{Inv}(k)} \prod_{i \in \text{Wex}(\sigma)} (\eta(\sigma, i) + 1).$$

Proof. In the proof of Theorem 2.1 of [22], replace U, D with $\mathfrak{N}, \mathfrak{P}$: in the conclusion, this replaces $\sum_{\text{rank } x=k} e(x)$ with $\sum_{|t|=k+1} m(\bullet; t)$. \square

For example, a sum over the four involutions 123, 213, 132, and 321 of Σ_3 gives

$$\sum_{|t|=4} m(\bullet; t) = 1 \cdot 2 \cdot 3 + 1 \cdot 3 + 1 \cdot 2 + 1 \cdot 1 = 12.$$

3. KREIMER'S HOPF ALGEBRA

In this section we discuss the Hopf algebra \mathcal{H}_K defined by D. Kreimer and his collaborators [1], [3], [13], [14], [15], [16] in connection with renormalization. As an algebra, \mathcal{H}_K is generated by rooted trees; so as a vector space, \mathcal{H}_K is generated by monomials in rooted trees, i.e., “forests” of rooted trees. For a rooted tree t , we give the corresponding generator degree $|t|$ in \mathcal{H}_K ; in degree 0, \mathcal{H}_K is generated by the unit element 1. For example, the degree-3 part of \mathcal{H}_K is generated as a vector space by the four elements



There is a linear map $B_+ : \mathcal{H}_K \rightarrow k\{\mathcal{T}\}$ which takes a forest to a single tree with a new root vertex connected to all the roots of the forest: e.g.,

$$B_+(\bullet \mid) = \text{diagram of a root connected to two children, one of which has a child of its own}.$$

The map B_+ takes the degree- n part of \mathcal{H}_K onto $k\{\mathcal{T}_n\}$: if we set $B_+(1) = \bullet$, then B_+ is a vector space isomorphism. We write B_- for the inverse of B_+ . On the other hand, except for the degree shift, \mathcal{H}_K is just the symmetric algebra on $k\{\mathcal{T}\}$. Thus, if $T_n = \dim k\{\mathcal{T}_n\} = |\mathcal{T}_n|$, we have

$$(5) \quad \sum_{n \geq 0} T_n x^n = \prod_{n \geq 1} \frac{1}{(1 - x^n)^{T_{n-1}}}$$

from which we can compute recursively $T_0 = 1, T_1 = 1, T_2 = 2, T_3 = 4, T_4 = 9$, etc. (see [19] for more information).

To define the bialgebra structure on \mathcal{H}_K , we let the counit send all elements of positive degree to 0, and the unit element 1 in degree 0 to $1 \in k$. The comultiplication Δ has $\Delta(1) = 1 \otimes 1$,

$$(6) \quad \Delta(t) = t \otimes 1 + (\text{id} \otimes B_+) \Delta(B_-(t))$$

for a rooted tree t , and $\Delta(t_1 t_2 \cdots t_n) = \Delta(t_1) \Delta(t_2) \cdots \Delta(t_n)$ for monomials $t_1 t_2 \cdots t_n$.

Equation (6) gives a recursive definition of the coproduct, but there is also a nonrecursive definition in terms of cuts. A cut of a rooted tree t is a set of edges of t . A cut is elementary if its cardinality is 1. When the elements of a cut C of t are removed, what remains is a collection of rooted trees: the one containing the root is denoted $R^C(t)$, and the remaining rooted trees form a monomial denoted $P^C(t)$. For example, if t is the tree



and C consists of the dotted edges, then

$$R^C(t) = \begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \bullet \end{array} \quad \text{and} \quad P^C(t) = \bullet \bullet.$$

The order of a cut C of t is the largest number of edges in C between the root of t and any of its terminal vertices: a cut of order at most 1 is called admissible. The empty cut \emptyset is the only cut of order 0 (note that $R^\emptyset(t) = t$ and $P^\emptyset(t) = 1$). The following formula for $\Delta(t)$ is proved in [3].

Proposition 3.1. *For a rooted tree t , $\Delta(t)$ can be written*

$$\Delta(t) = t \otimes 1 + \sum_{C \text{ admissible cut of } t} P^C(t) \otimes R^C(t).$$

We can define growth and pruning operators N and P on \mathcal{H}_K as follows. The growth operator N is simply \mathfrak{N} extended as a derivation, i.e.,

$$N(t_1 t_2 \cdots t_n) = \sum_{i=1}^n t_1 \cdots \mathfrak{N}(t_i) \cdots t_n.$$

We also define P as a derivation, but set $P(t) = \mathfrak{P}(t)$ only for $|t| \geq 2$; we put $P(\bullet) = 1$. If $D : \mathcal{H}_K \rightarrow \mathcal{H}_K$ is the extension of \mathfrak{D} as a derivation (i.e., the linear map that multiplies a monomial by its degree), then the identity

$$(7) \quad [P, N] = D$$

holds. To prove equation (7), note that both sides are derivations, and so it suffices to prove it for rooted trees t ; but in that case, (7) follows from Proposition 2.2. The map B_+ interacts with the growth and pruning operators as follows.

Proposition 3.2. *For monomials u of \mathcal{H}_K ,*

1. $B_+ P(u) = \mathfrak{P} B_+(u)$,
2. $B_+ N(u) = \mathfrak{N} B_+(u) - B_+(\bullet u)$.

Proof. Suppose $u = t_1 \cdots t_k$ with each $|t_i| \geq 1$. Then applying B_+ to

$$P(u) = P(t_1)t_2 \cdots t_k + t_1P(t_2)t_3 \cdots t_k + \cdots + t_1 \cdots t_{k-1}P(t_k)$$

gives a sum of rooted trees that includes all those obtained by removing terminal edges of $B_+(u)$, and the cases with $t_i = \bullet$ (hence $P(t_i) = 1$) work correctly: these are exactly those cases where an edge coming out of the root of $B_+(u)$ is terminal. If $u = 1$, then $\mathfrak{P}B_+(1) = \mathfrak{P}(\bullet) = 0 = B_+P(1)$. So in any case, $B_+P(u)$ coincides with $\mathfrak{P}B_+(u)$.

Now for $u = t_1 \cdots t_k$, B_+ applied to

$$N(u) = N(t_1)t_2 \cdots t_k + t_1N(t_2)t_3 \cdots t_k + \cdots + t_1 \cdots t_{k-1}N(t_k)$$

will include all those trees obtained by adding a new edge to each vertex of $B_+(u)$ except one—the “new” root vertex. Thus, $B_+N(u)$ is missing the term obtained by adding a new edge to the root of $B_+(u)$, namely $B_+(\bullet u)$. On the other hand, if $u = 1$ we have $B_+N(1) = 0 = \mathfrak{N}(\bullet) - B_+(\bullet)$. \square

We can extend the inner product of the previous section to \mathcal{H}_K by setting

$$(u_1, u_2) = (B_+(u_1), B_+(u_2))$$

for monomials u_1, u_2 ; there is no ambiguity since $(B_+(t), B_+(t')) = (t, t')$ for rooted trees t, t' . With this definition, we can state the adjointness relation between P and N .

Proposition 3.3. *On \mathcal{H}_K , the adjoint of P with respect to the inner product above is $N + M_\bullet$, where M_\bullet is the operator that sends u to $\bullet u$; equivalently, the adjoint of N is $P - \frac{\partial}{\partial \bullet}$.*

Proof. Let u_1, u_2 be monomials of \mathcal{H}_K . Then

$$(u_1, P(u_2)) = (B_+(u_1), B_+P(u_2)) = (B_+(u_1), \mathfrak{P}B_+(u_2)) = (\mathfrak{N}B_+(u_1), B_+(u_2)),$$

from which the first statement follows using the second part of Proposition 3.2. For the second statement, note that $\frac{\partial}{\partial \bullet}$ is adjoint to M_\bullet . \square

We now compute the characteristic polynomial of the restriction PN_k of PN to the degree- k part of \mathcal{H}_K . Let $\text{Ch}(L, \lambda) = \det(\lambda I - L)$ for a linear transformation L .

Proposition 3.4. *For $k \geq 1$,*

$$\text{Ch}(PN_k, \lambda) = \left(\lambda - \binom{k+1}{2} \right) \prod_{r=0}^{k-1} \left(\lambda - \sum_{j=0}^r (k-j) \right)^{T_{k-r} - T_{k-r-1}},$$

where as above T_i is the dimension of the degree- i part of \mathcal{H}_K .

Proof. We follow the proof of [21, Theorem 4.1]. Evidently PN_1 is the identity, and so the result holds for $k = 1$; assume it inductively for $k \geq 1$. From elementary linear algebra,

$$\text{Ch}(NP_{k+1}, \lambda) = \lambda^{T_{k+1} - T_k} \text{Ch}(PN_k, \lambda),$$

while from equation (7) we have

$$\text{Ch}(PN_{k+1}, \lambda) = \text{Ch}(NP_{k+1}, \lambda - (k+1))$$

since $D_{k+1} = (k+1)I$. The induction step then follows. \square

The preceding result implies that N_k is injective for all $k \geq 1$ and that P_k is surjective for $k \geq 2$; of course P_1 is also surjective. In addition, the maximal eigenvalue of PN_k is $\binom{k+1}{2}$. In fact, the element

$$f_k = N^{k-1}(\bullet) = \sum_{|t|=k} n(\bullet; t)t$$

is a corresponding eigenvector. To see this, note that

$$\begin{aligned} PN_k(f_k) &= P(f_{k+1}) = \sum_{|t'|=k} t' \sum_{|t|=k+1} n(\bullet; t)m(t'; t) \\ &= \sum_{|t'|=k} n(\bullet; t') \binom{k+1}{2} t' = \binom{k+1}{2} f_k, \end{aligned}$$

where we have used Proposition 2.8. The f_k are the “naturally grown forests” of [3] (where they are denoted δ_k).

The following result, which describes how N behaves with respect to the coproduct, is essentially [3, Prop. 6]. We give the proof since it can be stated concisely and illustrates the use of Proposition 3.1.

Proposition 3.5. $\Delta N = (N \otimes \text{id} + \text{id} \otimes N + M_\bullet \otimes D)\Delta$.

Proof. Since both sides are derivations, it suffices to show that

$$\Delta N(t) = (N \otimes \text{id} + \text{id} \otimes N + M_\bullet \otimes D)\Delta(t)$$

for any rooted tree t . As in the proof of Proposition 2.2, write $N(t) = \sum_i t_i$, where each t_i is the result of adding an edge to t . Then

$$\begin{aligned} \Delta N(t) &= \sum_i t_i \otimes 1 + \sum_i \sum_{C_i \text{ admissible cut of } t_i} P^{C_i}(t_i) \otimes R^{C_i}(t_i) \\ &= N(t) \otimes 1 + \sum_i \sum_{C_i \text{ admissible cut of } t_i} P^{C_i}(t_i) \otimes R^{C_i}(t_i). \end{aligned}$$

Now each cut C_i of t_i either includes the “new” edge or it does not. Suppose first that C_i does not include the new edge. Then C_i corresponds to a cut C of t and either $P^{C_i}(t_i) \otimes R^{C_i}(t_i)$ is a term in $P^C(t) \otimes NR^C(t)$ (if the new edge is in the component of the root) or a term in $NP^C(t) \otimes R^C(t)$ (if it is not). Together with the leading term $N(t) \otimes 1$, these give all the terms of $(N \otimes \text{id} + \text{id} \otimes N)\Delta(t)$.

Now suppose that C_i includes the new edge of t_i . If C is the cut of t given by C_i minus the new edge, then the new edge must have been attached to a vertex of $R^C(t)$ (by the definition of admissibility), and so

$$P^{C_i}(t_i) \otimes R^{C_i}(t_i) = \bullet P^C(t) \otimes R^C(t).$$

Since (for each admissible cut C of t) there are $|R^C(t)|$ vertices to which the new edge could be attached, terms of this form contribute $(M_\bullet \otimes D)\Delta(t)$. \square

Remark. It follows from this result that the f_k , $k \geq 1$, generate a sub-Hopf-algebra of \mathcal{H}_K . This Hopf algebra is isomorphic to the graded dual of the universal enveloping algebra of \mathcal{A}^1 , the Lie algebra of formal vector fields on \mathbf{R} that vanish to order 2 at the origin (see [3]).

Since \mathcal{H}_K is a locally finite commutative Hopf algebra, its graded dual \mathcal{H}_K^{gr} is a locally finite cocommutative Hopf algebra, hence (by the results of [17]) the universal enveloping algebra of the Lie algebra $\mathcal{P}(\mathcal{H}_K^{gr})$, the primitives of \mathcal{H}_K^{gr} . Primitives of \mathcal{H}_K^{gr} are dual to indecomposables of \mathcal{H}_K , and so are linear combinations of elements Z_t for rooted trees t , where $\langle Z_t, u \rangle = \delta_{t,u}$ for monomials $u \in \mathcal{H}_K$. The duals of N and P can be described as follows.

Proposition 3.6. 1. N^* is given by $N^*(Z_\bullet) = 0$,

$$(8) \quad N^*(Z_t) = \sum_{|t'|=|t|-1} n(t'; t) Z_{t'}$$

for $|t| \geq 2$, and

$$(9) \quad N^*(wv) = (N^*w)v + w(N^*v) + \frac{\partial w}{\partial Z_\bullet} |v|v$$

for $w, v \in \mathcal{H}_K^{gr}$.

2. $P^*(w) = Z_\bullet w$ for $w \in \mathcal{H}_K^{gr}$.

Proof. To prove the statements about $N^*(Z_t)$, note that $\langle N^*(Z_t), u \rangle = \langle Z_t, N(u) \rangle$ is zero unless u is a scalar multiple of t' , for some $t' \triangleleft t$; but then equation (8) follows from equation (1). Equation (9) follows from Proposition 3.5 since the multiplication in \mathcal{H}_K^{gr} is induced by Δ .

For the second part, let t be a rooted tree. If we write $P(t) = \sum_i t^{(i)}$ as in the proof of Proposition 2.2, then evidently

$$\bullet \otimes P(t) = \sum_i \bullet \otimes t^{(i)}$$

are (by Proposition 3.1) exactly those terms of $\Delta(t)$ of the form $\bullet \otimes t'$. Now let $u = t_1 t_2 \cdots t_n$ be a monomial of \mathcal{H}_K . Then

$$\begin{aligned} \Delta(u) &= \prod_{i=1}^n \Delta(t_i) \\ &= \prod_{i=1}^n (1 \otimes t_i + \bullet \otimes P(t_i) + \cdots) \\ &= 1 \otimes t_1 \cdots t_n + \bullet \otimes (P(t_1)t_2 \cdots t_n + \cdots + t_1 \cdots t_{n-1}P(t_n)) + \cdots \\ &= 1 \otimes u + \bullet \otimes P(u) + \cdots \end{aligned}$$

and thus

$$\langle Z_\bullet w, u \rangle = \langle Z_\bullet \otimes w, \Delta(u) \rangle = \langle w, P(u) \rangle = \langle P^*(w), u \rangle$$

for all $w \in \mathcal{H}_K^{gr}$ and monomials u of \mathcal{H}_K . □

Remark. The Lie algebra $\mathcal{P}(\mathcal{H}_K^{gr})$ is in fact free: see [5].

4. THE GROSSMAN-LARSON HOPF ALGEBRA

We can define a noncommutative multiplication on the graded vector space $k\{\mathcal{T}\}$ as follows. Let t, t' be rooted trees, and suppose $B_-(t) = t_1 t_2 \cdots t_k$. There are $|t'|^k$ rooted trees obtainable by attaching each of the k rooted trees t_1, t_2, \dots, t_k to some vertex of t' (by a new edge): let $t \circ t' \in k\{\mathcal{T}\}$ be the sum of these trees (if $t = \bullet$, we

define $t \circ t'$ to be t'). For example,

$$\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} \circ \begin{array}{c} \bullet \\ | \\ \bullet \end{array} = \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + 2 \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ | \\ \bullet \end{array} + \begin{array}{c} \bullet \\ | \\ \bullet \end{array}$$

while

$$\begin{array}{c} \bullet \\ | \\ \bullet \end{array} \circ \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} = \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array} + 2 \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ | \\ \bullet \end{array}.$$

This product makes $k\{\mathcal{T}\}$ a graded algebra: note that for $t \in k\{\mathcal{T}_n\}$ and $t' \in k\{\mathcal{T}_m\}$, we have $t \circ t' \in k\{\mathcal{T}_{n+m}\}$. The element $\bullet \in \mathcal{T}_0$ is a two-sided identity. Note also that

$$B_+(\bullet) \circ t = \begin{array}{c} \bullet \\ | \\ \bullet \end{array} \circ t = \mathfrak{N}(t)$$

for any rooted tree t .

Now define a coproduct $\Delta : k\{\mathcal{T}\} \rightarrow k\{\mathcal{T}\} \otimes k\{\mathcal{T}\}$ by

$$(10) \quad \Delta(t) = \sum_{I \cup J = \{1, \dots, k\}} B_+(t_I) \otimes B_+(t_J)$$

where $B_-(t) = t_1 \cdots t_k$ and the sum is over pairs (I, J) of (possibly empty) subsets I, J of $\{1, \dots, k\}$ such that $I \cup J = \{1, \dots, k\}$: t_I means the product of t_i for $i \in I$. The following result is proved in [10] and [9]: the main things to check are the associativity of the product \circ [10, Lemma 2.6] and the compatibility of the coproduct with \circ [10, Lemma 2.8].

Proposition 4.1. *The vector space $k\{\mathcal{T}\}$ with product \circ and coproduct Δ is a graded Hopf algebra \mathcal{H}_{GL} .*

Since the coproduct Δ is cocommutative, by results of [17] it follows that \mathcal{H}_{GL} is the universal enveloping algebra on its Lie algebra $\mathcal{P}(\mathcal{H}_{GL})$ of primitives. From equation (10), elements of the form $B_+(t)$, where t is a rooted tree, are primitive. We call such elements “primitive trees”: they are those rooted trees whose root has exactly one child. If we let \mathcal{PT} be the set of primitive trees (graded, like \mathcal{T} , by the number of non-root vertices), then we have the following result (for another proof see [10, Theorem 4.1]).

Proposition 4.2. *The vector space $k\{\mathcal{PT}\}$ generated by the primitive trees is $\mathcal{P}(\mathcal{H}_{GL})$.*

Proof. Since

$$B_+(t_1) \circ B_+(t_2) = B_+(t_1 t_2) + B_+(B_+(t_1) \circ t_2),$$

$k\{\mathcal{PT}\} \subseteq \mathcal{P}(\mathcal{H}_{GL})$ is a sub-Lie-algebra. Also, since B_+ is an isomorphism of $k\{\mathcal{T}_{n-1}\}$ onto $k\{\mathcal{PT}_n\}$, we have $\dim k\{\mathcal{PT}_n\} = T_{n-1}$. Then the Poincaré-Birkhoff-Witt theorem implies that the universal enveloping algebra of $k\{\mathcal{PT}\}$ has the same dimension in grade n as does the symmetric algebra on $k\{\mathcal{PT}\}$: but in view of equation (5), this dimension is $T_n = \dim(\mathcal{H}_{GL})_n$. Hence $k\{\mathcal{PT}\} = \mathcal{P}(\mathcal{H}_{GL})$. \square

Suppose t_1, t_2, t_3 are rooted trees so that $|t_1| + |t_2| = |t_3|$. If there is an elementary cut C of t_3 so that

(11)
$$P^C(t_3) = t_1 \quad \text{and} \quad R^C(t_3) = t_2,$$

let $m(t_1, t_2; t_3)$ be the number of distinct elementary cuts C of t_3 for which equations (11) hold: otherwise, set $m(t_1, t_2; t_3) = 0$. If (and only if) $m(t_1, t_2; t_3) \neq 0$, it is also true that t_3 can be obtained by attaching (via a new edge) the root vertex of t_1 to some vertex of t_2 : let $n(t_1, t_2; t_3)$ be the number of vertices of t_2 for which this is true. Evidently,

$$n(\bullet, t_2; t_3) = n(t_2; t_3) \quad \text{and} \quad m(\bullet, t_2; t_3) = m(t_2; t_3)$$

for trees $t_2 \triangleleft t_3$; so we have generalized the multiplicities of §2. (The reader is warned that $n(t_1, t_2; t_3)$ as used in [3] and [5] is our $m(t_1, t_2; t_3)$.) We now show how symmetry groups can be used to relate the two multiplicities.

Proposition 4.3. *For rooted trees t_1, t_2, t_3 with $|t_1| + |t_2| = |t_3|$,*

$$|SG(t_1)| |SG(t_2)| m(t_1, t_2; t_3) = n(t_1, t_2; t_3) |SG(t_3)|.$$

Proof. For any rooted tree t and $v \in V(t)$, let $\text{Fix}(t_v, t) \leq SG(t)$ be the subgroup of $SG(t)$ that holds t_v (the subtree of t with v as root) pointwise fixed. We can assume there is an elementary cut $C = \{e\}$ of t_3 so that equations (11) hold (otherwise both sides of the conclusion are zero). If e has source v and target w , then t_w is isomorphic to t_1 . Also, if $\text{Orb}(e, t_3)$ is the orbit of e under $SG(t_3)$, then

$$m(t_1, t_2; t_3) = |\text{Orb}(e, t_3)| = |SG(t_3) / \text{Fix}(t_v, t_3) \times SG(t_w)| = \frac{|SG(t_3)|}{|\text{Fix}(t_v, t_3)| |SG(t_1)|}.$$

On the other hand, since $R^C(t_3)$ is isomorphic to t_2 ,

$$n(t_1, t_2; t_3) = |\text{Orb}(v, R^C(t_3))| = |SG(R^C(t_3)) / \text{Fix}(v, R^C(t_3))| = \frac{|SG(t_2)|}{|\text{Fix}(v, R^C(t_3))|}.$$

Since there is an evident identification of $\text{Fix}(t_v, t_3)$ with $\text{Fix}(v, R^C(t_3))$, we have

$$\frac{|SG(t_1)|}{|SG(t_3)|} m(t_1, t_2; t_3) = \frac{n(t_1, t_2; t_3)}{|SG(t_2)|}$$

and the conclusion follows. □

We can now use the inner product on Kreimer’s Hopf algebra \mathcal{H}_K to define an isomorphism of \mathcal{H}_{GL} onto the graded dual of \mathcal{H}_K .

Proposition 4.4. *There is an isomorphism $\chi : \mathcal{H}_{GL} \rightarrow \mathcal{H}_K^{gr}$ defined by*

$$\langle \chi(t), u \rangle = (B_-(t), u) = (t, B_+(u))$$

for any rooted tree t and monomial u of \mathcal{H}_K .

Proof. Since \mathcal{H}_K is locally finite, it suffices to prove that χ is an injective homomorphism. We first show χ is a homomorphism, i.e., that

$$\begin{aligned} \langle \chi(t_1 \circ t_2), u \rangle &= \langle \chi(t_1) \otimes \chi(t_2), \Delta(u) \rangle \\ &= \sum_u \langle \chi(t_1), u' \rangle \langle \chi(t_2), u'' \rangle = \sum_u (t_1, B_+(u')) (t_2, B_+(u'')) \end{aligned}$$

for any monomial u of \mathcal{H}_K with coproduct

$$(12) \quad \Delta(u) = \sum_u u' \otimes u''.$$

In view of Proposition 4.2, \mathcal{H}_{GL} is generated as an algebra by the primitive trees. So it suffices to show that

$$(13) \quad \langle \chi(B_+(t) \circ t_2), u \rangle = \sum_u (B_+(t), B_+(u'))(t_2, B_+(u'')) = \sum_u (t, u')(t_2, B_+(u'')).$$

Now from the definition of $n(t_1, t_2; t_3)$,

$$\begin{aligned} \langle \chi(B_+(t) \circ t_2), u \rangle &= (B_+(t) \circ t_2, B_+(u)) \\ &= \sum_{|t_3|=|t|+|t_2|} n(t, t_2; t_3)(t_3, B_+(u)) = n(t, t_2; B_+(u))|SG(B_+(u))|. \end{aligned}$$

On the other hand, if $\Delta(u)$ is given by equation (12), then

$$(14) \quad \Delta(B_+(u)) = B_+(u) \otimes 1 + \sum_u u' \otimes B_+(u'')$$

by equation (6). Now the only nonzero terms of

$$\sum_u (t, u')(t_2, B_+(u''))$$

are those with $u' = t$ and $t_2 = B_+(u'')$: and (comparing Proposition 3.1 with equation (14)) there are $m(t, t_2; B_+(u))$ such terms. Hence

$$\sum_u (t, u')(t_2, B_+(u'')) = m(t, t_2; B_+(u))|SG(t)||SG(t_2)|,$$

and equation (13) follows from Proposition 4.3: thus, χ is a homomorphism.

Now suppose $v = \sum_i a_i t_i \in \ker \chi$. Then

$$\langle \chi(v), u \rangle = \sum_i a_i (t_i, B_+(u)) = 0$$

for all monomials u of \mathcal{H}_K . But setting $u = B_-(t_i)$ implies that $a_i = 0$ for each i ; so $v = 0$. \square

Remark. In [18, Prop. 2.1] (and also in [9, Theorem 14.16]) it is wrongly asserted that the map sending $B_+(t)$ to Z_t induces an isomorphism of \mathcal{H}_{GL} onto \mathcal{H}_K^{gr} : the error is due to a failure to distinguish the multiplicities $n(t_1, t_2; t_3)$ and $m(t_1, t_2; t_3)$, since Panaite confuses the coefficients in

$$[Z_{t_1}, Z_{t_2}] = \sum_{|t_3|=|t_1|+|t_2|} (m(t_1, t_2; t_3) - m(t_2, t_1; t_3))Z_{t_3}$$

with those in

$$B_+(t_1) \circ B_+(t_2) - B_+(t_2) \circ B_+(t_1) = \sum_{|t_3|=|t_1|+|t_2|} (n(t_1, t_2; t_3) - n(t_2, t_1; t_3))B_+(t_3).$$

In fact, since

$$\langle \chi(B_+(t)), u \rangle = (t, u) = |SG(t)|\delta_{t,u} = |SG(t)|\langle Z_t, u \rangle,$$

we have $\chi(B_+(t)) = |SG(t)|Z_t$.

We can use the isomorphism χ to express the duals of P and N as maps of \mathcal{H}_{GL} (cf. Proposition 3.6 above).

Proposition 4.5. 1. The map $\chi^{-1}P^*\chi : \mathcal{H}_{GL} \rightarrow \mathcal{H}_{GL}$ is left multiplication by $B_+(\bullet)$, i.e., $\chi^{-1}P^*\chi(t) = \mathfrak{N}(t) = B_+(\bullet) \circ t$.

2. For rooted trees t , $\chi^{-1}N^*\chi(t) = \mathfrak{P}(t) - B_+ \frac{\partial}{\partial \bullet} B_-(t)$.

Proof. Using Propositions 2.3, 3.2, and 3.3, we have for any rooted tree t and monomial u of \mathcal{H}_K ,

$$\langle \chi(t), P(u) \rangle = (t, B_+P(u)) = (t, \mathfrak{P}B_+(u)) = (\mathfrak{N}(t), B_+(u)) = \langle \chi(\mathfrak{N}(t)), u \rangle$$

and

$$\begin{aligned} \langle \chi(t), N(u) \rangle &= (t, B_+N(u)) = (t, \mathfrak{N}B_+(u)) - (t, B_+(\bullet u)) \\ &= (\mathfrak{P}(t), B_+(u)) - \left(\frac{\partial}{\partial \bullet} B_-(t), u \right) = \left\langle \chi \left(\mathfrak{P}(t) - B_+ \frac{\partial}{\partial \bullet} B_-(t) \right), u \right\rangle. \end{aligned}$$

□

REFERENCES

1. D. J. Broadhurst and D. Kreimer, Renormalization automated by Hopf algebra, *J. Symbolic Comput.* **27** (1999), 581-600. MR **2000h**:81167
2. C. Brouder, Runge-Kutta methods and renormalization, *European Phys. J. C* **12** (2000), 521-534.
3. A. Connes and D. Kreimer, Hopf algebras, renormalization, and noncommutative geometry, *Comm. Math. Phys.* **199** (1998), 203-242. MR **99h**:81137
4. A. Connes and H. Moscovici, Hopf algebras, cyclic cohomology and the transverse index theorem, *Comm. Math. Phys.* **198** (1998), 199-246. MR **99m**:58186
5. L. Foissy, Finite-dimensional comodules over the Hopf algebra of rooted trees, *J. Algebra* **255** (2002), 89-120.
6. S. V. Fomin, The generalized Robinson-Schensted-Knuth correspondence (Russian), *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **155** (1986), Differentsialnaya Geometriya, Gruppy Li i Mekh. VIII, 156-175, 195; translation in *J. Soviet Math.* **41** (1988), 979-991. MR **88b**:06003
7. S. Fomin, Duality of graded graphs, *J. Algebraic Combin.* **3** (1994), 357-404. MR **95i**:05088
8. S. Fomin, Schensted algorithms for dual graded graphs, *J. Algebraic Combin.* **4** (1995), 5-45. MR **95m**:05246
9. J. M. Gracia-Bondía, J. C. Várilly, and H. Figueroa, *Elements of Noncommutative Geometry*, Birkhäuser, Boston, 2001. MR **2001h**:58038
10. R. Grossman and R. G. Larson, Hopf-algebraic structure of families of trees, *J. Algebra* **126** (1989), 184-210. MR **90j**:16022
11. M. E. Hoffman, An analogue of covering space theory for ranked posets, *Electron. J. Combin.* **8(1)** (2001), #R32. MR **2003d**:06004
12. D. E. Knuth, *The Art of Computer Programming*, vol. 3, 2nd ed., Addison-Wesley, Reading, MA, 1998. MR **56**:4281
13. D. Kreimer, On the Hopf algebra structure of perturbative quantum field theory, *Adv. Theor. Math. Phys.* **2** (1998), 303-334. MR **99e**:81156
14. D. Kreimer, On overlapping divergences, *Comm. Math. Phys.* **204** (1999), 669-689. MR **2000f**:81128
15. D. Kreimer, Chen's iterated integral represents the operator product expansion, *Adv. Theor. Math. Phys.* **3** (1999), 627-670. MR **2003b**:81128
16. D. Kreimer, *Knots and Feynman Diagrams*, Cambridge University Press, Cambridge, 2000. MR **2002f**:81078
17. J. W. Milnor and J. C. Moore, On the structure of Hopf algebras, *Ann. of Math. (2)* **81** (1965), 211-261. MR **30**:4259
18. F. Panaite, Relating the Connes-Kreimer and Grossman-Larson Hopf algebras built on rooted trees, *Lett. Math. Phys.* **51** (2000), 211-219. MR **2002a**:81177

19. N. J. A. Sloane, Online Encyclopedia of Integer Sequences, Sequence A000081, www.research.att.com/~njas/sequences/.
20. R. P. Stanley, *Ordered Structures and Partitions*, Memoirs Amer. Math. Soc. No. 119, American Mathematical Society, Providence, RI, 1972. MR **48**:10836
21. R. P. Stanley, Differential posets, *J. Amer. Math. Soc.* **1** (1988), 919-961. MR **89h**:06005
22. R. P. Stanley, Variations on differential posets, *Invariant Theory and Tableaux (Minneapolis, MN 1988)*, IMA Volumes in Mathematics and its Applications 19, Springer-Verlag, New York, 1990, pp. 145-165. MR **91h**:06004

DEPARTMENT OF MATHEMATICS, UNITED STATES NAVAL ACADEMY, ANNAPOLIS, MARYLAND 21402

E-mail address: meh@usna.edu

CONNECTIONS WITH PRESCRIBED FIRST PONTRJAGIN FORM

MAHUYA DATTA

ABSTRACT. Let P be a principal $O(n)$ bundle over a C^∞ manifold M of dimension m . If $n \geq 5m + 4 + 4\binom{m+1}{4}$, then we prove that every differential 4-form representing the first Pontrjagin class of P is the Pontrjagin form of some connection on P .

1. INTRODUCTION

Let P be a principal $O(n)$ bundle over a C^∞ manifold M of dimension m , and let $p_i \in H^{4i}(M)$ denote the i -dimensional Pontrjagin class of P . We address the question whether a $4i$ -form representing the class p_i is a Pontrjagin form of some connection on P . In [1] we considered the top-dimensional Pontrjagin class p_d of a principal $O(n)$ bundle P over a $4d$ -dimensional *open* manifold M for $n \geq 2d$, and we gave a homotopy classification of connections α on P that satisfy $p_d(\alpha) = \omega$, where ω is a volume form on M . In this paper, we take up the case of the first Pontrjagin form and prove the following result.

Theorem 1.1. *If $n \geq 5m + 4 + 4\binom{m+1}{4}$, then every differential 4-form representing the first Pontrjagin class p_1 is the Pontrjagin form of some connection on P . Moreover, when M is a closed manifold, the same is true for $n \geq 5m + 4\binom{m}{4}$.*

Here $\binom{m}{k}$ denotes the integer $\frac{m!}{k!(m-k)!}$.

We observe that when $n > m$, then P reduces to the direct sum $P_1 \oplus P_2$ of two principal bundles, where P_1 is an $O(m)$ bundle and P_2 is the trivial $O(n-m)$ bundle on M . Since the Pontrjagin form is additive, the above observation reduces the problem to finding a connection on a trivial principal bundle with a given exact form as its Pontrjagin form.

Now, if an exact 4-form on M can be expressed as the sum of q primary monomials of the form $df_1 \wedge df_2 \wedge df_3 \wedge df_4$, where the f_i 's are smooth functions on M , then we can explicitly construct a connection on the trivial principal $O(2q)$ -bundle over M by taking a 2×2 block

$$\alpha = \begin{pmatrix} 0 & f_1 df_2 - f_3 df_4 \\ -f_1 df_2 + f_3 df_4 & 0 \end{pmatrix}$$

along the principal diagonal for each such monomial. It can be seen easily that the Pontrjagin form of such a connection is the given exact form on M . Indeed, we can prove the following result (compare ([2], 3.4.1 (B'))).

Received by the editors September 26, 2002 and, in revised form, February 14, 2003.

2000 *Mathematics Subject Classification.* Primary 53C05, 53C23, 58J99.

Key words and phrases. Principal bundle, connections, first Pontrjagin form.

Theorem 1.2. *Every exact 4-form $d\omega$ on M can be expressed as the sum of q primary monomials for $q \geq 2(m+1) + 2\binom{m+1}{4}$. Furthermore, if M is a closed manifold, then the same is true for $q \geq 2m + 2\binom{m}{4}$.*

In view of the above discussion it is easy to see that Theorem 1.1 follows from Theorem 1.2.

We employ the sheaf-theoretic and analytic techniques of the theory of the h -principle [2] to prove the above result. We observe that an exact 4-form $d\omega$ can be expressed as the sum of q primary monomials if and only if there is a map $f : M \rightarrow \mathbb{R}^{4q}$ such that

$$d\omega = \sum_{i=1}^q f_i^* \sigma,$$

where σ is the canonical volume form on \mathbb{R}^4 and $f_i : M \rightarrow \mathbb{R}^4$, $i = 1, 2, \dots, q$, are components of f . The maps characterized by the above equation are solutions to a certain first-order partial differential equation. The associated partial differential operator being infinitesimally invertible on an open subset, we apply Gromov's formulation of the Implicit Function Theorem in the infinite-dimensional setup to make way for the sheaf techniques.

In Section 2, following Gromov [2], we briefly describe the notion of infinitesimal inversion of partial differential operators and state some results relating to the solution sheaf of infinitesimally invertible operators. We shall assume that the reader is familiar with the language of the h -principle, in particular with the terms: partial differential relations, holonomic section, the h -principle, (micro)flexible sheaf and sharply moving diffeotopy. For a brief review of terminology and sheaf techniques in the h -principle we refer to the Appendix of [1]. In section 3 we consider immersions in a manifold N with a fixed k -form σ and prove the h -principle for " σ -regular" immersions that induce a given k -form on the source manifold. This has been shown by observing that the relevant differential operator is infinitesimally invertible on the space of " σ -regular" immersions. In section 4 we prove that if (N, σ) is the q -fold product of the k -dimensional Euclidean space with canonical volume form, then σ -regular immersions are generic for q sufficiently large. Using genericity of σ -regular maps, we then prove the second part of Theorem 1.2. Finally, by applying the h -principle of σ -regular maps (Section 3), we prove the full form of Theorem 1.2 and the main result of this paper.

2. INFINITESIMAL INVERSION OF DIFFERENTIAL OPERATORS

Let $X \rightarrow M$ be a C^∞ fibration and $G \rightarrow M$ be a C^∞ vector bundle over a manifold M . We denote by \mathcal{X}^α and \mathcal{G}^α respectively the spaces of C^α sections of X and G with the fine C^α topology, for $\alpha = 1, 2, \dots, \infty$. Let $\mathcal{D} : \mathcal{X}^r \rightarrow \mathcal{G}^0$ be a C^∞ differential operator of order r , so that if x is a $C^{\alpha+r}$ section of X , then $\mathcal{D}(x)$ is a C^α section of G for $\alpha = 1, 2, \dots, \infty$.

Let $T_{\text{vert}}(X) \subset TX$ denote the subspace of vertical vectors (i.e., tangent to the fibres of the fibration $X \rightarrow M$) in the tangent bundle TX of X . For a section $x : M \rightarrow X$, we denote the induced vector bundle $x^*T_{\text{vert}}(X)$ by Y_x . When x is C^α , this bundle is C^β -smooth for $\beta \leq \alpha$ and we denote by \mathcal{Y}_x^β the space of C^β sections of this induced bundle. The space \mathcal{Y}_x^α can be realized as the infinite-dimensional tangent space of \mathcal{X}^α at x . We define the linearization L_x of the operator

\mathcal{D} at x as follows:

$$L_x : \mathcal{Y}_x^r \longrightarrow \mathcal{G}^0,$$

$$L_x(y) = \frac{d}{dt} \mathcal{D}(x_t)|_{t=0},$$

where $\{x_t : t \geq 0\}$ is a 1-parameter family of sections of X with $x_0 = x$ and $\frac{dx_t}{dt}|_{t=0} = y$. Clearly, L_x is a linear differential operator of order r in y and $L(x, y) = L_x(y)$ is a differential operator of order r in both x and y .

Let $A \subset X^{(d)}$ be an open subset of the d -jet space of sections of X for some $d \geq r$. Following Gromov, we shall call such a subset an open *differential relation* of order d . A solution of A will also be referred to as an A -regular section of X . Let \mathcal{A} denote the space of solutions of the relation A . Clearly, \mathcal{A} is contained in \mathcal{X}^d , and $\mathcal{A}^{\alpha+d} = \mathcal{A} \cap \mathcal{X}^{\alpha+d}$ is an open subset of $\mathcal{X}^{\alpha+d}$ in the fine $C^{\alpha+d}$ topology.

\mathcal{D} is said to be *infinitesimally invertible* over the subset $\mathcal{A} \subset \mathcal{X}$ if for every $x \in \mathcal{A}$ there is a linear differential operator $M_x : \mathcal{G}^s \longrightarrow \mathcal{Y}_x^0$ of a certain order s (independent of x) such that the following properties are satisfied:

- (1) The global operator

$$M : \mathcal{A}^d \times \mathcal{G}^s \longrightarrow T(\mathcal{X}^0)$$

is a differential operator that is given by a C^∞ map $A \oplus G^{(s)} \longrightarrow T_{vert}(X)$, where $T(\mathcal{X}^0)$ denotes the tangent bundle of \mathcal{X}^0 .

- (2) $L(x, M(x, g)) = g$ for all $x \in \mathcal{A}^{d+r}$ and $g \in \mathcal{G}^{r+s}$. In other words, M_x is a right inverse of L_x .

The integer d is called the *defect* of the infinitesimal inversion M ([2], 2.3.1).

We now state an infinite-dimensional Implicit Function Theorem due to Gromov which generalizes Nash's theory in the context of differential operators.

Let \mathcal{D} be a C^∞ differential operator of order r . Suppose \mathcal{D} admits an infinitesimal inversion of order s and of defect d . Let us fix a Riemannian metric on M . Let $\alpha > \max(d, 2r + s)$.

Theorem 2.1 ([2], 2.3.2). *For every $x \in \mathcal{A}^\infty$ there exists a fine $C^{\alpha+s}$ neighbourhood \mathcal{B}_x of the zero section in the space $\mathcal{G}^{\alpha+s}$ and an operator $\mathcal{D}_x^{-1} : \mathcal{B}_x \longrightarrow \mathcal{A}^\alpha$ such that*

- (1) $\mathcal{D}_x^{-1}(0) = x$.
- (2) (*Inversion property*) $\mathcal{D}(\mathcal{D}_x^{-1}(g)) = \mathcal{D}(x) + g$.
- (3) If $g \in \mathcal{B}_x$ is $C^{\beta+s}$ -smooth, for $\beta \geq \alpha$, then $\mathcal{D}_x^{-1}(g)$ is C^β -smooth.
- (4) (*Locality*) The value of $\mathcal{D}_x^{-1}(g)$ at any point $v \in M$ does not depend on the behaviour of x and g outside the unit ball $B_v(1)$ in M with centre v relative to the fixed metric on M .

In particular, the operator $\mathcal{D} : \mathcal{A}^\infty \longrightarrow \mathcal{G}^\infty$ is an open map in the respective fine C^∞ topologies.

It is to be noted that the local inverse \mathcal{D}_x^{-1} depends on the Riemannian metric on M . If we choose an appropriate Riemannian metric on M , then applying the locality property of the inverse in Theorem 2.1 we can prove

Proposition 2.2 ([2], 2.3.2). *If \mathcal{D} is infinitesimally invertible, then the sheaf of A -regular solutions of the differential equation $\mathcal{D}(x) = g$ is microflexible.*

We now consider some partial differential relations which have the same C^∞ solutions, namely the solutions to the equation $\mathcal{D}(x) = g$. Let $\mathcal{R}^\alpha \subset X^{(\alpha+r)}$

consist of $(\alpha + r)$ -jets of infinitesimal solutions of $\mathcal{D} = g$ of order α and let \mathcal{R}^0 be denoted as \mathcal{R} . Recall that x is an infinitesimal solution of $\mathcal{D} = g$ of order α if $\mathcal{D}(x) - g$ has zero α -jet. Define

$$\mathcal{R}_\alpha = \mathcal{R}^\alpha \cap (p_d^{\alpha+r})^{-1}(A),$$

where $p_d^{\alpha+r} : X^{(\alpha+r)} \longrightarrow X^{(d)}$ is the canonical projection map for $\alpha \geq d - r$. The relations \mathcal{R}_α have the same C^∞ solutions for all $\alpha \geq d - r$, namely the C^∞ solutions of the equation $\mathcal{D}(x) = g$ in \mathcal{A} (such a solution, from now on, will be referred to as an A -regular solution of the equation).

Let Φ denote the sheaf of A -regular solutions of the equation $\mathcal{D}(x) = g$ with the C^∞ compact open topology and let Ψ_α be the sheaf of sections of \mathcal{R}_α with C^0 compact open topology. It is a consequence of Theorem 2.1 that

Proposition 2.3 ([2], 2.3.2). *If $\alpha \geq \max(d + s, 2r + 2s)$, then an infinitesimal solution of \mathcal{R}_α can be deformed to a local solution. Furthermore, the map $J : \Phi \longrightarrow \Psi_\alpha$, defined by $J(\phi) = j_\phi^{r+\alpha}$, is a local weak homotopy equivalence. In other words, \mathcal{R}_α satisfies the local h -principle.*

3. THE h -PRINCIPLE OF ISOMETRIC σ -REGULAR MAPS

We start with the following definition.

Definition 3.1 ([2], 3.4.1). *Let (N, σ) be a smooth manifold with a closed k -form σ . A smooth map $f : M \longrightarrow N$ is said to be σ -regular if for each $x \in M$, the map*

$$\begin{array}{ccc} I_\sigma : T_{f(x)}N & \longrightarrow & \Lambda^{k-1}(T_x M), \\ \partial & \longmapsto & f^*(\partial \cdot \sigma) \end{array}$$

is surjective for all $x \in M$.

A σ -regular map is necessarily an immersion.

Let ω be a given k -form on M for $k \geq 2$. We call a map $f : (M, \omega) \longrightarrow (N, \sigma)$ *isometric* if $f^*\sigma = \omega$. In this section we shall prove the h -principle for σ -regular isometric maps $(M, \omega) \longrightarrow (N, \sigma)$ in the following situation:

- (1) both σ and ω are exact;
- (2) $M = M_0 \times \mathbb{R}$;
- (3) ω is induced from a k -form on M_0 by the projection map $p : M_0 \times \mathbb{R} \longrightarrow M_0$.

Let $\mathcal{D} : C^\infty(M, N) \longrightarrow \Omega^k(M)$ denote the first-order differential operator on the space of C^∞ maps $f : M \longrightarrow N$ with values in the space of k -forms $\Omega^k(M)$ defined by $\mathcal{D}(f) = f^*\sigma$. Since σ is a closed form, the sheaf of solutions of $\mathcal{D} = \omega$ is not microflexible ([2], 3.4.1). Now, suppose that $\sigma = d\sigma_1$ and $\omega = d\omega_1$ for some $(k - 1)$ -forms σ_1 and ω_1 on N and M respectively. If f is a smooth immersion such that $f^*\sigma = \omega$, then locally on any contractible set the above equation reduces to $f^*\sigma_1 + d\phi = \omega_1$ for some $(k - 2)$ -form ϕ on M . Conversely, if (f, ϕ) is a pair satisfying $f^*\sigma_1 + d\phi = \omega_1$, then $f^*\sigma = \omega$. Let

$$\bar{\mathcal{D}} : C^\infty(M, N) \times \Omega^{k-2}(M) \longrightarrow \Omega^{k-1}(M)$$

denote the differential operator defined by $\bar{\mathcal{D}}(f, \phi) = f^*\sigma_1 + d\phi$, where $f : M \longrightarrow N$ is a smooth map and ϕ is a differential $(k - 2)$ -form on M . Note that the pairs (f, ϕ) can be realized as sections of the fibre bundle $(M \times N) \oplus \Lambda^{k-2}(M)$ over M which will be denoted by E for future reference.

The linearization $L_{(f,\phi)}$ of the operator $\bar{\mathcal{D}}$ at (f,ϕ) can be obtained as follows: Consider a smooth 1-parameter family of sections $\{(f_t, \phi_t)\}$ in E such that $(f_0, \phi_0) = (f, \phi)$. Then

$$L_{(f,\phi)}(\partial, \tilde{\phi}) = \frac{d}{dt} \mathcal{D}(f_t, \phi_t)|_{t=0},$$

where $\partial = \frac{df_t}{dt}|_{t=0}$ and $\frac{d\phi_t}{dt}|_{t=0} = \tilde{\phi}$. Hence,

$$L_{(f,\phi)}(\partial, \tilde{\phi}) = f^* d(\partial \cdot \sigma_1) + f^*(\partial \cdot d\sigma_1) + d\tilde{\phi},$$

where ∂ is a vector field on N along f and $\tilde{\phi}$ is a $(k-2)$ -form on M . The equation $L_{(f,\phi)} = \omega_1$ can be solved for $(\partial, \tilde{\phi})$ if the following system has a solution:

$$\begin{aligned} f^*(\partial \cdot d\sigma_1) &= \omega_1, \\ f^*(\partial \cdot \sigma_1) + \tilde{\phi} &= 0. \end{aligned}$$

Now the above system of equations is solvable for $(\partial, \tilde{\phi})$ if f is a σ -regular map. Thus the operator $\bar{\mathcal{D}}$ is infinitesimally invertible on all those (f, ϕ) for which f is σ -regular ([2]). Since σ -regularity is an open condition and depends only on the first jet of a map, the space of pairs (f, ϕ) for which f is σ -regular corresponds to the solution space of an open differential relation $A \subset E^{(1)}$, where $E^{(1)}$ is the 1-jet bundle of sections of the fibre bundle E mentioned above. Hence the operator $\bar{\mathcal{D}}$ has the zeroth-order inversion (i.e., $s = 0$, where s is defined as in Section 2) with defect $d = 1$.

Let $\bar{\Phi}$ be the sheaf of σ -regular solutions of the equation $\mathcal{D}(f) = \omega$ and let $\bar{\Phi}$ be the sheaf of pairs (f, ϕ) satisfying the equation $\bar{\mathcal{D}} = \omega_1$ where f is σ -regular. There is a canonical map $\bar{\Phi} \rightarrow \bar{\Phi}$ that takes a pair (f, ϕ) onto f . Furthermore, $\bar{\Phi}(x)$ has the same homotopy type as the space $\bar{\Phi}(x)$.

Let $\bar{\mathcal{R}}^\alpha \subset E^{(\alpha+1)}$ consist of $(\alpha+1)$ -jets of infinitesimal solutions of $\bar{\mathcal{D}} = \omega_1$ of order α and let $\bar{\mathcal{R}}_\alpha = \bar{\mathcal{R}}^\alpha \cap (p_1^{\alpha+1})^{-1}(A)$, where $p_1^{\alpha+1} : E^{(\alpha+1)} \rightarrow E^{(1)}$ is the canonical projection. The following proposition is a direct consequence of Proposition 2.2 and Proposition 2.3.

Proposition 3.2. (i) The solution sheaf $\bar{\Phi}$ of $\bar{\mathcal{D}} = \omega_1$ is microflexible.

(ii) The 3-jet map $j^3 : \bar{\Phi}(x) \rightarrow \bar{\Psi}(x)$ is a weak homotopy equivalence for every $x \in M$, where $\bar{\Psi}$ is the sheaf of sections of $\bar{\mathcal{R}}_2$. In particular, if (f, ϕ) is an infinitesimal solution of order 2 of $\bar{\mathcal{D}} = \omega_1$ where f is also σ -regular, then (f, ϕ) can be homotoped to a local solution of the equation.

Theorem 3.3. Let σ be an exact k -form on N as above and let $M = M_0 \times \mathbb{R}$. If the form $\omega = d\omega_1$ on M is induced from an exact k -form on M_0 by the projection map $p : M_0 \times \mathbb{R} \rightarrow M_0$, then every section of $\bar{\mathcal{R}}_2$ is homotopic to a holonomic section (in the space of continuous sections of $\bar{\mathcal{R}}_2$ with C^0 compact open topology).

Proof. Let $\bar{\Psi}$ denote the sheaf of sections of the jet bundle $E^{(3)}$ with images in $\bar{\mathcal{R}}_2$. We shall prove that

$$j^3 : \bar{\Phi}|_{M_0} \rightarrow \bar{\Psi}|_{M_0}$$

is a weak homotopy equivalence.

First observe that the fibre-preserving diffeomorphisms of $M_0 \times \mathbb{R}$ act on the sheaf $\bar{\Phi}$. To see this take a smooth immersion $f : M_0 \times \mathbb{R} \rightarrow N$ and a $(k-2)$ -form ϕ such that $f^*\sigma_1 + d\phi = \omega_1$, where $\omega_1 = p^*\omega_0$ for some $(k-1)$ -form on M_0 . Let

$\delta : M_0 \times \mathbb{R} \longrightarrow M_0 \times \mathbb{R}$ be a fibre-preserving diffeomorphism so that $p \circ \delta = p$. Define the action of δ by

$$\delta.(f, \phi) \mapsto (f \circ \delta, \delta^* \phi).$$

Then,

$$(f \circ \delta)^* \sigma_1 + d(\delta^* \phi) = \delta^*(f^* \sigma_1 + d\phi) = \delta^* \omega_1 = \delta^* p^* \omega_0 = p^* \omega_0 = \omega_1.$$

Also, if f is σ -regular, then so is $f \circ \sigma$.

On the other hand, the fibre-preserving diffeotopies sharply move M_0 in $M_0 \times \mathbb{R}$ ([2], [1]). Since the sheaf $\bar{\Phi}$ is microflexible (Proposition 3.2), we conclude that the restriction of $\bar{\Phi}$ to M_0 is flexible ([2], 2.3.2, [1]).

A standard argument proves that the sheaf $\bar{\Psi}$ is flexible ([2], 1.4.2 (A')) and Proposition 3.2 (ii) says that

$$j^3 : \bar{\Phi}(x) \longrightarrow \bar{\Psi}(x)$$

is a weak homotopy equivalence for every $x \in M$. Then by the Sheaf Homomorphism Theorem ([2], 2.2.1 (B))

$$j^3 : \bar{\Phi}|_{M_0} \longrightarrow \bar{\Psi}|_{M_0}$$

is a weak homotopy equivalence.

Finally, the theorem follows from the observation that $M_0 \times \mathbb{R}$ can be deformed into an arbitrary small neighbourhood of M_0 by means of fibre-preserving diffeomorphisms of M . \square

Let \mathcal{R}_1 consist of 2-jets of σ -regular infinitesimal solutions of order 1 of the equation $\mathcal{D} = \omega$ and let $\Gamma(\mathcal{R}_1)$ denote the space of continuous sections of \mathcal{R}_1 with C^0 compact open topology. Then we have the following.

Corollary 3.4. *An arbitrary section of $\mathcal{R}_1 \subset J^2(M, N)$ is homotopic to a holonomic section in $\Gamma(\mathcal{R}_1)$. Hence, the σ -regular isometric C^∞ immersions $f : (M_0 \times \mathbb{R}, d\omega_1 = d\omega_0 \oplus 0) \longrightarrow (N, d\sigma_1)$ satisfy the h -principle. Furthermore, if M is an open manifold, then σ -regular isotropic immersions satisfy the h -principle.*

Proof. Let (f, ϕ) be a second-order infinitesimal solution at x of the equation $\bar{\mathcal{D}} = \omega_1$. Then $j_{(f^* \sigma_1 - d\phi)}^2 = j_{\omega_1}^2$ at x . There is a bundle map $\Delta_{k-1} : (\Lambda^{k-1}(M))^{(2)} \longrightarrow \Lambda^k(M)^{(1)}$ associated to the exterior differential operator d such that $\Delta_{k-1}(j_\tau^2) = j_{d\tau}^1$. Then applying Δ_{k-1} on the preceding equation we get $j_{f^* \sigma}^1(x) = j_\omega^1(x)$. Thus f is an infinitesimal solution of order 1 of the equation $\mathcal{D} = \omega$. Hence we have the canonical map $p : \mathcal{R}_2 \longrightarrow \mathcal{R}_1$ that maps $(j_f^3(x), j_\phi^3(x))$ onto $j_f^2(x)$. We shall prove that this map is surjective and that fibres of p are affine subspaces. This would imply that p has a section, and then the first part of the corollary would follow from the above theorem.

To prove that p is surjective, consider the following sequence of vector bundles:

$$\dots \longrightarrow (\Lambda^{k-2}(M))^{(3)} \xrightarrow{\Delta_{k-2}} (\Lambda^{k-1}(M))^{(2)} \xrightarrow{\Delta_{k-1}} (\Lambda^k(M))^{(1)} \longrightarrow \dots$$

where the bundle maps Δ_k are induced by the exterior differential operator d as $\Delta_k \circ j_\tau^i = j_{d\tau}^{i-1}$. By the formal Poincaré Lemma this sequence is exact.

Let f be a first-order infinitesimal solution of $\mathcal{D} = \omega$ at $x \in M$, which is also σ -regular, so that $j_f^2(x) \in \mathcal{R}_1$. Then $j_{f^* \sigma}^1(x) = j_\omega^1(x)$ and consequently $j_{f^* \sigma_1}^2(x) - j_{\omega_1}^2(x)$ is in $\ker \Delta_{k-1}$. Hence there exists a 3-jet $j_\phi^3(x)$ such that

$$j_{f^* \sigma_1}^2(x) - j_{\omega_1}^2(x) = \Delta_{k-2}(j_\phi^3(x)) = j_{d\phi}^2(x).$$

Therefore, $(j_f^3(x), j_\phi^3(x)) \in \bar{\mathcal{R}}_2$ and p is surjective.

Now let $j_f^2(x) \in \mathcal{R}_1$. Then $p^{-1}(j_f^2(x))$ consists of all pairs $(j_g^3(x), j_\phi^3(x)) \in E^{(3)}$ such that $j_g^2(x) = j_f^2(x)$ and $j_{d\phi}^2(x) = j_{g^*\sigma_1}^2(x) - j_{\omega_1}^2(x)$, equivalently, $j_\phi^3(x) \in \Delta_{k-2}^{-1}(j_{g^*\sigma_1}^2(x) - j_{\omega_1}^2(x))$. This shows that the fibres of p are affine subspaces and that $p : \bar{\mathcal{R}}_2 \longrightarrow \mathcal{R}_1$ is an affine bundle. This proves the first part of the corollary.

To prove the second part, one has to note, in addition, that the zero form is invariant under any diffeomorphism of M , and M can be deformed into an arbitrary small neighbourhood of its $(m-1)$ -skeleton by an isotopy. \square

4. EXISTENCE OF σ -REGULAR IMMERSIONS INDUCING ω

Let σ_0 be a closed k -form on a manifold N_0 , and let N be the q -fold Cartesian product of N_0 with the k -form $\sigma = \sum_{i=1}^q \pi_i^* \sigma_0$, where $\pi_i : N \longrightarrow N_0$ is the projection onto the i -th factor. We first determine when the σ -regular maps exist generically and then prove the existence of isometric maps, applying the results obtained in the previous sections.

Definition 4.1. An immersion $f = (f_1, f_2, \dots, f_q) : M \longrightarrow N$ is said to be σ_0 -large if $f_1^* \sigma_0, \dots, f_q^* \sigma_0$ span the k -th exterior bundle $\Lambda^k(M)$; this means, for every k -form ω on M , there exist continuous functions $\beta_i : M \longrightarrow \mathbb{R}$, $i = 1, \dots, q$, such that

$$\omega = \sum_{i=1}^q \beta_i f_i^* \sigma_0.$$

Let

$$\tilde{\mathcal{A}} = \{(\ell_1, \dots, \ell_q) \in J_x^1(M, N) : \ell_1^* \sigma_0, \dots, \ell_q^* \sigma_0 \text{ span } \Lambda^k(T_x M), x \in M\}.$$

If $f = (f_1, \dots, f_q)$ is a solution of $\tilde{\mathcal{A}}$, then $f_1^* \sigma_0(x), \dots, f_q^* \sigma_0(x)$ span $\Lambda^k(T_x M)$ for each $x \in M$. Moreover, it follows from the lemma below that the σ_0 -large maps are precisely the solutions of the relation $\tilde{\mathcal{A}}$.

Lemma 4.2. Let $\omega_1, \dots, \omega_q$ be k -forms on M such that for each $x \in M$, $\omega_1(x), \dots, \omega_q(x)$ span $\Lambda^k(T_x M)$. Then $\omega_1, \dots, \omega_q$ span the space of k -forms $\Omega^k(M)$ over the ring of continuous functions on M .

Since $\tilde{\mathcal{A}}$ is an open relation, the σ_0 -large immersions form an open set in the fine C^∞ topology. Next we observe that

Proposition 4.3. If $f = (f_1, \dots, f_q) : M \longrightarrow N$ is a σ_0 -large immersion, then f is σ -regular.

Proof. Let $f = (f_1, \dots, f_q)$ be a σ_0 -large immersion of M into the q -fold product of N_0 . If $\partial_1, \partial_2, \dots, \partial_q$ are vector fields on M , then we have the relation

$$\sum_{i=1}^q \partial_i \cdot f_i^* \sigma_0 = \sum_{i=1}^q f_i^* (\bar{\partial}_i \cdot \sigma_0) = f^* ((\bar{\partial}_1, \dots, \bar{\partial}_q) \cdot \sigma),$$

where $\bar{\partial}_i = (f_i)_* \partial_i$ is a vector field on N along f_i . The proposition now follows from the following simple observation.

Lemma 4.4. *If $\omega_1, \dots, \omega_q$ are linear k -forms on \mathbb{R}^m spanning $\Lambda^k(\mathbb{R}^m)$, then the linear map*

$$\begin{aligned} \mathbb{R}^m \times \dots \times \mathbb{R}^m &\longrightarrow \Lambda^k(\mathbb{R}^{m-1}), \\ \partial_1, \dots, \partial_q &\longmapsto \sum_{i=1}^q \partial_i \cdot \omega_i \end{aligned}$$

is surjective.

In the rest of this article, M and (N, σ) will be as follows:

- (1) M will denote a manifold of dimension m ;
- (2) N will denote the q -fold Cartesian product of the Euclidean space \mathbb{R}^k ;
- (3) σ will denote the k -form obtained by summing the q canonical volume forms $\sigma_k := dy_1 \wedge \dots \wedge dy_k$ on each \mathbb{R}^k factor, where y_1, y_2, \dots, y_k are the canonical coordinates on \mathbb{R}^k .

Proposition 4.5. *If $q \geq m + \binom{m}{k}$, then $f_1^* \sigma_k, \dots, f_q^* \sigma_k$ span the k -th exterior bundle of M for generic $(f_1, \dots, f_q) : M \rightarrow N$. Consequently, if $q \geq 2m + 2\binom{m}{k}$, there exists a σ_k -large immersion $f : M \rightarrow (N, \sigma)$ such that $f^*(\sigma) = 0$.*

Proof. Here $N = \mathbb{R}^{qk}$ and $\sigma = \bigoplus_{i=1}^q \sigma_k$. Fix a basis e_1, e_2, \dots, e_m for \mathbb{R}^m . Let L be a linear map from \mathbb{R}^m to \mathbb{R}^{qk} . Then L can be expressed as $L = (L_1, L_2, \dots, L_q)$, where L_i is the projection of L onto the i -th copy of \mathbb{R}^k .

If L is σ_k -large, then the forms $L_1^* \sigma_k, L_2^* \sigma_k, \dots, L_q^* \sigma_k$ span the bundle $\Lambda^k(\mathbb{R}^m)$. Note that the $k \times k$ cofactors of L_i correspond to the values of $L_i^* \sigma_k$ on the k -tuples of basis vectors $(e_{i_1}, \dots, e_{i_k})$, where $\{i_1, i_2, \dots, i_k\}$ is an ordered subset of $\{1, 2, \dots, m\}$. If \bar{L}_i denotes the column vector formed by the $k \times k$ cofactors of the matrix L_i , then by a σ -large condition on L is meant that $\bar{L} = (\bar{L}_1, \dots, \bar{L}_q)$ has the maximum rank. Let Σ' consist of all linear maps $L = (L_1, \dots, L_q) : \mathbb{R}^m \rightarrow \mathbb{R}^{qk}$ such that $\text{rank } \bar{L}$ is strictly less than $l = \binom{m}{k}$; in other words, any $l \times l$ cofactor of \bar{L} is zero. Therefore, Σ' is semialgebraic and hence stratified ([2], 1.3.1). Moreover, the codimension of Σ' in $L(\mathbb{R}^m, \mathbb{R}^{qk})$ is $q - \binom{m}{k} + 1$.

Let Σ be the subset of the 1-jet space $J^1(M, N)$ consisting of all 1-jets $j_f^1(x)$ such that $\{f_i^* \sigma_k : i = 1, 2, \dots, q\}$ do not span $\Lambda_x^k(M)$. Hence a map $f : M \rightarrow N$ is σ_k -large if its 1-jet map misses the set Σ . Since σ has global symmetry, the singular set Σ in the 1-jet space fibres over M and therefore it is stratified with codimension $q - \binom{m}{k} + 1$. Hence by the Thom Transversality Theorem, a generic map is σ_k -large if $q - \binom{m}{k} \geq m$.

Now, let $f = (f_1, \dots, f_q) : M \rightarrow \mathbb{R}^{qk}$ be a σ_k -large immersion; then define $\bar{f} = (\bar{f}_1, \dots, \bar{f}_q)$ as follows:

$$\bar{f}_i = (f_{i2}, f_{i1}, f_{i3}, \dots, f_{ik}),$$

where $f_i = (f_{i1}, f_{i2}, f_{i3}, \dots, f_{ik}) : M \rightarrow \mathbb{R}^k$. Note that $\bar{f}_i^* \sigma_k = -f_i^* \sigma_k$ for every i . Hence $(f, \bar{f}) : M \rightarrow \mathbb{R}^{qk} \times \mathbb{R}^{qk}$ is a σ_k -large immersion of M into \mathbb{R}^{2qk} that pulls back $\sigma \oplus \sigma$ onto the zero form on M . \square

Theorem 4.6. *Let M be a closed manifold. If $q \geq 2m + 2\binom{m}{k}$, then every exact form on M can be induced from σ by a σ -regular immersion $f : M \rightarrow N$. Consequently, every exact k -form on a closed m -dimensional manifold is expressible as the sum of q primary monomials for $q \geq 2m + 2\binom{m}{k}$.*

Proof. Let $\tau = y_1 dy_2 \wedge \cdots \wedge dy_k$ so that $\sigma_k = d\tau$. It follows from Section 3 and Proposition 4.5 that the operator

$$\bar{\mathcal{D}} : (f_1, f_2, \dots, f_q, \phi) \mapsto \sum_{i=1}^q f_i^* \tau + d\phi$$

is infinitesimally invertible on σ_k -large immersions which exist generically for $q \geq m + \binom{m}{k}$. Hence by Theorem 2.1, the image of σ_k -large immersions under $\bar{\mathcal{D}}$ is a nonempty open set in the fine C^∞ topology for $q \geq m + \binom{m}{k}$. Moreover, when $q \geq 2m + 2\binom{m}{k}$, there exists a σ_k -large immersion $f = (f_1, \dots, f_q) : M \rightarrow \mathbb{R}^{kq}$ such that $\sum_{i=1}^q f_i^* \sigma_k = 0$, which implies that $\sum_{i=1}^q f_i^* \tau$ is a closed form. As a consequence, $\text{Image } \bar{\mathcal{D}}$ contains a closed $(k-1)$ -form c .

Let M now be closed and let $\omega = d\alpha$ be exact. Then for sufficiently small $\lambda > 0$, $c + \lambda\alpha \in \text{Image } \bar{\mathcal{D}}$. In other words, there exists a σ_k -large immersion (g_1, g_2, \dots, g_q) and a $(k-2)$ -form ψ such that

$$c + \lambda\alpha = \sum_{i=1}^q g_i^* \tau + d\psi,$$

and therefore

$$\omega = \sum_{i=1}^q \left(\frac{1}{k\sqrt{\lambda}} g_i \right)^* \sigma_k.$$

Clearly, $(\frac{1}{k\sqrt{\lambda}} g_1, \frac{1}{k\sqrt{\lambda}} g_2, \dots, \frac{1}{k\sqrt{\lambda}} g_q)$ is σ -regular and this completes the proof of the theorem. \square

The next result is an immediate consequence of the above theorem.

Corollary 4.7. *If M is arbitrary, then every compactly supported exact k -form on M can be induced by a σ -regular immersion $f : M \rightarrow (N, \sigma)$ for $q \geq 2m + 2\binom{m}{k}$.*

Corollary 4.8. *If P is a principal $O(n)$ bundle over a closed manifold M , then every compactly supported 4-form on M representing the first Pontrjagin class of P is the Pontrjagin form of some connection on P , for $n \geq 5m + 4\binom{m}{4}$.*

The proof of the above corollary will be similar to that of Corollary 4.10 and we omit it here.

Theorem 4.9 ([2], 3.4.1 (B')). *Let (N, σ) be the q -fold product of (\mathbb{R}^k, σ_k) for $k \geq 2$. If $q \geq 2(m+1) + 2\binom{m+1}{k}$, then an arbitrary exact k -form on M can be induced by a σ -regular immersion $f : M \rightarrow N$. Therefore, every exact k -form on an m -dimensional manifold is expressible as the sum of q primary monomials for $q \geq 2(m+1) + 2\binom{m+1}{k}$.*

Proof. Let x_1, x_2, \dots, x_m denote a local coordinate system on M . Then a k -form ω on M can be represented locally as:

$$\omega = \sum_I \omega_I dx_I,$$

where I runs over all multi-indices (i_1, i_2, \dots, i_k) for $1 \leq i_1 < i_2 < \cdots < i_k \leq m$, $\omega_I = \omega_{i_1, i_2, \dots, i_k}$ are smooth functions defined locally on M , and $dx_I = dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$.

Recall that $\sigma = \bigoplus_{i=1}^q \sigma_k$, where σ_k is the canonical volume form on \mathbb{R}^k . If $f = (f_1, \dots, f_q) : M \longrightarrow \mathbb{R}^{qk}$ is a smooth map, then $f^*\sigma = \omega$ defines for each multi-index $I = (i_1, i_2, \dots, i_k)$ with $1 \leq i_1 < i_2 < \dots < i_k \leq m$ an equation E_I :

$$\begin{aligned} & \sum_{j=1}^q \det \left(\frac{\partial f_j}{\partial x_{i_1}} \quad \frac{\partial f_j}{\partial x_{i_2}} \quad \dots \quad \frac{\partial f_j}{\partial x_{i_k}} \right) \\ &= \sum_j \sum_{\alpha} (-1)^{\text{sgn } \alpha} \frac{\partial f_{j\alpha_1}}{\partial x_{i_1}} \frac{\partial f_{j\alpha_2}}{\partial x_{i_2}} \dots \frac{\partial f_{j\alpha_k}}{\partial x_{i_k}} = \omega_I, \end{aligned}$$

where α represents an element of the symmetry group S_k on k letters $\{1, 2, \dots, k\}$, and $\{f_{j\alpha}\}$ denote the components of f_j .

Differentiating E_I with respect to x_p , $p \in \{1, 2, \dots, m\}$, we get an equation E_I^p :

$$\sum_{j=1}^q \sum_{\pi, \alpha} (-1)^{\text{sgn } \alpha} \frac{\partial^2 f_{j\alpha_1}}{\partial x_p \partial x_{i_{\pi(1)}}} \frac{\partial f_{j\alpha_2}}{\partial x_{i_{\pi(2)}}} \dots \frac{\partial f_{j\alpha_k}}{\partial x_{i_{\pi(k)}}} = \frac{\partial \omega_I}{\partial x_p},$$

where π is an element of the symmetry group S_k on k letters $\{1, 2, \dots, k\}$.

The collection $\{f_j(x), \frac{\partial f_j}{\partial x_i}(x), \frac{\partial^2 f_j}{\partial x_i \partial x_p}(x)\}$ defines the 2-jet of the function $f : M \longrightarrow \mathbb{R}^{qk}$ at x . If f satisfies the equation $f^*\sigma = \omega$, then its 2-jet map satisfies the above system of equations.

Replacing the partial derivatives in the above equations by ordinary variables, namely substituting

$$\frac{\partial f_{j\alpha}}{\partial x_i} = v_i^{j\alpha}, \quad \frac{\partial^2 f_{j\alpha}}{\partial x_i \partial x_p} = v_{ip}^{j\alpha},$$

we obtain a system of equations $\{\bar{E}_I, \bar{E}_I^p\}$, where I runs over all multi-indices (i_1, i_2, \dots, i_k) with $1 \leq i_1 < i_2 < \dots < i_k \leq m$. It can be verified that this system of equations is independent of coordinate transformation and defines the relation \mathcal{R}^1 in the 2-jet space.

Note that $\{\bar{E}_I^p\}$ is a system of $m \binom{m}{k}$ equations that are linear in the variables $v_{ip}^{j\alpha}$, the total number of which is $kqm(m+1)/2$. Let A denote the coefficient matrix of the vector $\{v_{ip}^{j\alpha}\}$ in the system $\{\bar{E}_I^p\}$. The system of equations $\{\bar{E}_I^p\}$ has a solution if the matrix A has the maximum rank everywhere. Since $k \geq 2$, the condition "rank $A < \text{maximum}$ " defines a stratified subset Σ in the 1-jet space $J^1(M, \mathbb{R}^{qk})$. If q is such that $kq(m+1)/2 \geq \binom{m}{k} + 1$, then $\text{codim } \Sigma > m$ and hence by the Thom Transversality Theorem, j_f^1 misses Σ for generic f . In other words, we get a map f for which the following system of equations has a solution for each $x \in M$:

$$(1) \quad \sum_{j=1}^q \sum_{\pi, \alpha} (-1)^{\text{sgn } \alpha} \frac{\partial f_{j\alpha_2}}{\partial x_{i_{\pi(2)}}} \dots \frac{\partial f_{j\alpha_k}}{\partial x_{i_{\pi(k)}}} v_{p i_{\pi(1)}}^{j\alpha_1} = \frac{\partial \omega_I}{\partial x_p}.$$

Moreover, the space of solutions is an affine subspace in \mathbb{R}^d of codimension $m \binom{m}{k}$, where $d = kqm(m+1)/2$. Therefore, if $q \geq \binom{m+1}{k} + m + 1$, then there exists a map $f : M \times \mathbb{R} \longrightarrow (N, \sigma)$ that is a σ_k -large immersion and for which the system of equations (1) has a solution, say $v_{ip}^{j\alpha} = \bar{v}_{ip}^{j\alpha}$. Since f is σ_k -large, there exist continuous real-valued functions β_i on M such that $\omega = \sum_{i=1}^q \beta_i L_i^* \sigma_k$, where

L_i denotes the derivative map df_i . Define for each $i = 1, 2, \dots, q$, a bundle map $\bar{L}_i : TM \longrightarrow T\mathbb{R}^k$ by

$$\bar{L}_i(x) = \begin{cases} df_i(x) & \text{if } \beta_i(x) > 1, \\ d\bar{f}_i(x) & \text{if } \beta_i(x) < 1, \end{cases}$$

where \bar{f}_i is obtained from f_i by interchanging the first two component functions. Take $T = (L_1, \dots, L_q, \bar{\beta}_1 \bar{L}_1, \dots, \bar{\beta}_q \bar{L}_q)$, where $\bar{\beta}_i = |\beta_i - 1|^{\frac{1}{k}}$. Note that T_i extends continuously over all of M if we define it to be identically zero on the set $\beta_i^{-1}(1)$. Thus we get a σ -regular bundle map $T : TM \longrightarrow T\mathbb{R}^{2q}$ such that $T^*\sigma = \omega$. We extend this (locally) to a section of \mathcal{R}_1 by taking $v_{ip}^{j\alpha} = \bar{v}_{ip}^{j\alpha}$ for $j \leq q$ and $v_{ip}^{j\alpha} = 0$ for $j > q$. These local solutions finally define a global section of \mathcal{R}_1 if we patch them together by a partition of unity. (Note that the system of equations (1) is linear in $v_{ip}^{j\alpha}$.) We now conclude the existence of an isometric immersion by Theorem 3.3. \square

Theorems 4.9 and 4.6 prove Theorem 1.2.

Corollary 4.10. *Let P be a principal $O(n)$ bundle over a manifold M of dimension m , and let $n \geq 5m + 4 + 4\binom{m+1}{4}$. Then every 4-form on M representing the first Pontrjagin class of P is the Pontrjagin form of some connection on P .*

Proof. If $n > \dim M$, then P can be reduced to $P_1 \oplus P_2$, where P_1 is a principal $O(m)$ bundle and P_2 is the trivial $O(n - m)$ bundle over M . This may be seen easily if we view a principal $O(n)$ bundle as a frame bundle associated to some vector bundle of rank n . Moreover, we have a canonical inclusion $Q = P_1 \oplus P_2 \xrightarrow{i} P$ that takes the fibres of $P_1 \oplus P_2$ canonically into the fibres of P . Now we prove that the Pontrjagin forms of the bundles P and Q are the same. It is a standard fact that a connection α_Q on Q can be extended uniquely to a connection α_P on P such that $i^*\alpha_P = \alpha_Q$. We shall show that $p_1(\alpha_Q) = p_1(\alpha_P)$. We recall that the first Pontrjagin form $p_1(\alpha_Q)$ is uniquely determined by the equation

$$(2) \quad \pi_Q^* p_1(\alpha_Q) = \text{trace}(D\alpha_Q \wedge D\alpha_Q),$$

where D stands for the covariant differentiation and π_Q denotes the projection map $Q \longrightarrow M$. Similarly, $\pi_P^* p_1(\alpha_P) = \text{trace}(D\alpha_P \wedge D\alpha_P)$ ([3]). Taking pull-back by i we get $i^*\pi_P^* p_1(\alpha_P) = \text{trace}(D\alpha_Q \wedge D\alpha_Q)$. Since $\pi_P \circ i = \pi_Q$, the left-hand side is equal to $\pi_Q^* p_1(\alpha_P)$. Hence by equation (2) and the uniqueness property, $p_1(\alpha_P) = p_1(\alpha_Q)$. Moreover, the Pontrjagin form is additive, so that if α_1 and α_2 are connections on P_1 and P_2 , respectively, then $p_1(\alpha_1 \oplus \alpha_2) = p_1(\alpha_1) + p_1(\alpha_2)$. In view of the above observation it is enough to show that every exact form on M is the Pontrjagin form of some connection on the trivial principal $O(n)$ bundle for $n \geq 4(m + 1) + 4\binom{m+1}{4}$.

Let $d\omega$ be an exact 4-form on M . We have proved in Theorem 4.9 that an exact 4-form on a manifold of dimension m can be expressed as the sum of q primary monomials for $q \geq 2(m + 1) + 2\binom{m+1}{4}$. Let $d\omega = 2 \sum_{i=1}^q df_{i1} \wedge df_{i2} \wedge df_{i3} \wedge df_{i4}$, where f_{ij} are smooth functions on M and where q satisfies the above relation. Now consider an $\mathfrak{o}(2q)$ -valued 1-form α on M such that corresponding to each monomial $df_{i1} \wedge df_{i2} \wedge df_{i3} \wedge df_{i4}$ there exists a 2×2 block

$$\alpha_i = \begin{pmatrix} 0 & f_{i1} df_{i2} - f_{i3} df_{i4} \\ -f_{i1} df_{i2} + f_{i3} df_{i4} & 0 \end{pmatrix}$$

along the principal diagonal, all other elements being zero. Clearly α is a connection on the trivial principal $O(2q)$ -bundle over M and its first Pontrjagin form is

$$\begin{aligned} p_1(\alpha) &= \sum_{i=1}^q p_1(\alpha_i) \\ &= \sum_{i=1}^q \text{trace}(D\alpha_i \wedge D\alpha_i) \\ &= \sum_{i=1}^q \text{trace}(d\alpha_i \wedge d\alpha_i) = d\omega. \end{aligned}$$

This completes the proof. □

Corollaries 4.8 and 4.10 prove Theorem 1.2.

REFERENCES

1. M. Datta, A note on Pontrjagin forms, *Proceedings of the American Mathematical Society* **128** (2000), pp. 3723 - 3729. MR **2001k**:53142
2. M. Gromov, *Partial Differential Relations*, *Ergebnisse der Mathematik und ihrer Grenzgebiete* 3. Folge Band 9. Springer-Verlag, Berlin, 1986. MR **90a**:58201
3. S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vols. I, II, Interscience Tracts in Pure and Applied Mathematics, John Wiley and Sons, 1963, 1969. MR **27**:2945; MR **38**:6501

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF CALCUTTA, 35 P. BARUA SARANI, CALCUTTA 700019, INDIA

E-mail address: mahuyad@hotmail.com

SELF-INTERSECTION CLASS FOR SINGULARITIES AND ITS APPLICATION TO FOLD MAPS

TORU OHMOTO, OSAMU SAEKI, AND KAZUHIRO SAKUMA

Dedicated to Professor Takuo Fukuda on the occasion of his 60th birthday

ABSTRACT. Let $f : M \rightarrow N$ be a generic smooth map with corank one singularities between manifolds, and let $S(f)$ be the singular point set of f . We define the self-intersection class $I(S(f)) \in H^*(M; \mathbf{Z})$ of $S(f)$ using an incident class introduced by Rimányi but with twisted coefficients, and give a formula for $I(S(f))$ in terms of characteristic classes of the manifolds. We then apply the formula to the existence problem of fold maps.

1. INTRODUCTION

Given a smooth map $g : M \rightarrow N$ between smooth manifolds, does there exist a smooth map homotopic to g that has at most “nice” singularities? If not, then what is the obstruction?

Embeddings and immersions, which have no singularities at all, have been particularly well studied since the concept of a manifold was established around 1930. For example, Whitney proved that every n -dimensional manifold can be embedded in \mathbf{R}^{2n} and immersed in \mathbf{R}^{2n-1} . Furthermore, the Smale–Hirsch theory [18] gives a satisfactory answer to the existence problem of immersions in terms of homotopy theory. Note that these results make sense only when $n = \dim M < \dim N = p$.

On the other hand, similar problems for the case $n \geq p$ seem to have been hardly studied except for maps of open manifolds (for example, see [22]). If M is closed and $n \geq p$, then every smooth map $f : M \rightarrow N$ into an open manifold N must have singularities. In this paper, we consider *fold maps* — smooth maps that have at most fold singularities (for details, see §4) — for such cases, and discuss their (non)existence problem. Note that a fold singularity is the simplest of the generic singularities: for example, a fold map for $p = 1$ and $N = \mathbf{R}$ is nothing but a Morse function.

When $p = 2$, the existence problem of fold maps has been solved: a smooth map of a closed connected n -dimensional manifold M with $n \geq 2$ into an orientable

Received by the editors September 12, 2002 and, in revised form, March 24, 2003.

2000 *Mathematics Subject Classification*. Primary 57R45; Secondary 57R42.

Key words and phrases. Self-intersection class, incident class, Thom polynomial, Pontrjagin class, twisted coefficient, fold map.

The first author has been partially supported by Grant-in-Aid for Scientific Research (No. 12740046), the Ministry of Education, Science and Culture, Japan. The second and the third authors have been partially supported by Grant-in-Aid for Scientific Research (No. 13640076), the Ministry of Education, Science and Culture, Japan. The third author has also been partially supported by Grant for Encouragement of Young Researchers, Kinki Univ. (G008).

surface is homotopic to a fold map if and only if the Euler characteristic of M is even [31], [20], [9]. Recall that generic maps into surfaces have fold and cusp singularities, and that the number of cusp singularities has the same parity as the Euler characteristic (or equivalently, the top Stiefel-Whitney class $w_n(M) \in H^n(M; \mathbf{Z}_2) = \mathbf{Z}_2$) of the source manifold. The above result shows that this is the only obstruction for the existence of fold maps into orientable surfaces. In fact, our problem is closely related to the concept of the Thom polynomial of a given singularity type, as explained below.

Let Σ be a singularity type, which means a certain union of orbits of the group $\text{Diff}(\mathbf{R}^n, 0) \times \text{Diff}(\mathbf{R}^p, 0)$ with respect to the right-left action on some jet space $J^s(n, p)$. For a generic map $f : M \rightarrow N$ between smooth manifolds, $\Sigma(f) \subset M$ denotes the singular point set of f of type Σ . It is known that for Σ good enough, the closure $\overline{\Sigma(f)}$ of $\Sigma(f)$ represents a homology class and that its Poincaré dual, denoted by $\text{Tp}(\Sigma)(f)$, can be described as a polynomial in characteristic classes of TM and f^*TN . This is called the *Thom polynomial* for Σ and does not depend on a particular choice of f (for example, see [17], [3], [24], [12]). Clearly, if $\text{Tp}(\Sigma)(f)$ or $\text{Tp}(\Sigma')(f)$ for some Σ' adjacent to Σ (i.e., $\Sigma' \subset \overline{\Sigma} \setminus \Sigma$) does not vanish, then there is no generic map \tilde{f} homotopic to f such that $\Sigma(\tilde{f}) = \emptyset$. However, in general, even if $\text{Tp}(\Sigma)(f)$ vanishes, we cannot conclude that the corresponding singularity can be eliminated by homotopy. An example in low dimensions was discovered by the second author, as follows.

Theorem 1.1 (Saeki [25], [26]). *Let M be a smooth 4-manifold with $H_*(M; \mathbf{Z}) \cong H_*(\mathbf{CP}^2; \mathbf{Z})$. Then there exists no fold map of M into any orientable 3-manifold.*

Namely, for M as above, every generic map $f : M \rightarrow N$ into an orientable 3-manifold N must have cusp singularities. Note that the Thom polynomials for cusp singularities (or A_2 -type singularities) and its adjacent swallowtail singularities (or A_3 -type singularities) both vanish in our case.

In the proof of Theorem 1.1, the following congruence, which the third author [29] proved by using a Rohlin type theorem, played an essential role: for a generic map $f : M \rightarrow N$ of a closed oriented 4-manifold M with $H_1(M; \mathbf{Z}) = 0$ into an orientable 3-manifold N , we have

$$(1.1) \quad S(f) \cdot S(f) \equiv 3\sigma(M) \pmod{4},$$

where $S(f) \subset M$ is the set of all singular points of f , called the *singular set* of f , $S(f) \cdot S(f)$ is the self-intersection number of $S(f)$ in M , and $\sigma(M)$ is the signature of M . In a sense, the congruence (1.1) gives us more information than the usual Thom polynomials.

The purpose of this paper is to describe what is the essential point behind the congruence (1.1) and to give its integral lift for general dimensions. For a certain generic map $f : M \rightarrow N$ between smooth manifolds whose singular point set $S(f)$ is a smooth submanifold, we define the *self-intersection class* of $S(f)$, denoted by $I(S(f))$, as the cohomology class in $H^*(M; \mathbf{Z})$ Poincaré dual to the homology class represented by the transverse intersection of $S(f)$ and its small perturbation in M . This class coincides with the Gysin map image of a special kind of the *incident class* introduced by Rimányi in his theory of Thom polynomials [24] but refined by using twisted coefficients (for more details, see §3). By using the desingularization

method, we obtain a formula for $I(S(f))$ as follows, where Σ^ℓ denotes the submanifold of the 1-jet bundle $J^1(M, N)$ consisting of the 1-jets of kernel dimension ℓ .

Theorem 1.2. *Let M and N be manifolds of dimensions n and p respectively, where M is closed and $n - p + 1 = 2k$, $n \geq p \geq 1$, $k \geq 1$. Furthermore, let $f : M \rightarrow N$ be a smooth map such that $j^1 f$ is transverse to Σ^{2k} and $\Sigma^\ell(f) = \emptyset$ for all $\ell \geq 2k + 1$. Then we have*

$$(1.2) \quad I(S(f)) \equiv p_k(TM - f^*TN) \in H^{4k}(M; \mathbf{Z}) \quad (\text{modulo } 2\text{-torsion}),$$

where $p_k(TM - f^*TN)$ is the k -th Pontrjagin class of the difference bundle $TM - f^*TN$.

In the terminology of [15, Chapter VI, §1], the above condition on $j^1 f$ is equivalent to f being 1-generic and having corank at most one everywhere (the *corank* of f at x is defined to be $\min(n, p) - \text{rank } df_x$). In this case, we say simply that f is a generic smooth map with corank one singularities. Note that then the singular set $S(f) = (j^1 f)^{-1}(\Sigma^{2k})$ is a smooth regular submanifold of M . Note also that this condition is generic provided that $n > 2p - 4$.

The above theorem implies, in particular, that for a generic map $f : M \rightarrow N$ of a closed oriented 4-manifold M into an orientable 3-manifold N , we have

$$(1.3) \quad S(f) \cdot S(f) = p_1[M] = 3\sigma(M) \in \mathbf{Z},$$

by the Hirzebruch signature formula, where $p_1[M]$ denotes the first Pontrjagin number of M . This is nothing but an integral lift of the congruence (1.1). Clearly, Theorem 1.1 can be proved by using (1.3). As another application of the formula (1.2), we have the following necessary condition for the existence of fold maps.

Theorem 1.3. *Let $g : M \rightarrow N$ be a smooth map between smooth manifolds, where M is closed. We assume that $n = \dim M$ and $p = \dim N$ satisfy $n - p + 1 = 2k$ for some positive odd integer k . If there exists a fold map homotopic to g , then there exists a cohomology class $x \in H^{2k}(M; \mathcal{O}_{TM-g^*TN})$ such that*

$$x \smile x \equiv p_k(TM - g^*TN) \in H^{4k}(M; \mathbf{Z}) \quad (\text{modulo } 4\text{-torsion}),$$

where \smile denotes the cup product and \mathcal{O}_{TM-g^*TN} is the orientation local system associated with the difference bundle $TM - g^*TN$.

The above theorem is very useful for obtaining nonexistence results for fold maps. For example, we will show that for certain dimension pairs (n, p) , every n -dimensional oriented cobordism class contains a connected manifold that admits no fold maps into certain p -dimensional manifolds (for details, see §4).

Throughout the paper, we work in the smooth category; that is, all manifolds and maps are differentiable of class C^∞ unless otherwise stated.

The authors would like to express their sincere gratitude to Julius Korbaš and Peter Zvengrowski for their useful suggestions.

2. GYSIN MAP WITH TWISTED COEFFICIENTS AND SELF-INTERSECTION CLASS

In this section, we recall the definition and some properties of the Gysin map with twisted coefficients induced by a smooth map between manifolds, and define the self-intersection class of a submanifold.

In the following, for a manifold X (or a vector bundle ξ), \mathcal{O}_X (resp. \mathcal{O}_ξ) will denote the orientation local system associated with X (resp. ξ).

Let $f : M \rightarrow N$ be a smooth map between smooth manifolds of dimensions n and p respectively, and \mathcal{L} an arbitrary local system over N . Take an embedding $\varepsilon : M \rightarrow \text{Int } D^r$ for some r and identify M with the image of the embedding $(f, \varepsilon) : M \rightarrow N \times D^r$. Let U be a tubular neighbourhood of M in $N \times D^r$ with projection $p_U : U \rightarrow M$, where U is identified with the total space of the normal bundle ν of M in $N \times D^r$. Note that the orientation local system \mathcal{O}_ν is isomorphic to that associated with the difference bundle $f^*TN - TM$, i.e., $\mathcal{O}_\nu \cong \mathcal{O}_{f^*TN - TM} \cong \mathcal{O}_M \otimes f^*\mathcal{O}_N$.

Then, we define the (twisted) *Gysin map* (or the *Umkehr map*)

$$f_! : H^*(M; f^*\mathcal{L} \otimes \mathcal{O}_{f^*TN - TM}) \rightarrow H^{*+(p-n)}(N; \mathcal{L})$$

induced by f by the composition

$$\begin{aligned} H^*(M; f^*\mathcal{L} \otimes \mathcal{O}_{f^*TN - TM}) &\xrightarrow{(p_U)^*} H^*(U; (p_U)^*f^*\mathcal{L} \otimes (p_U)^*\mathcal{O}_{f^*TN - TM}) \\ &\xrightarrow{\smile u} H^{*+(p+r-n)}(U, U \setminus M; (p_U)^*f^*\mathcal{L}) \\ &\xrightarrow{\cong} H^{*+(p+r-n)}(U, U \setminus M; (p_U)^*f^*\mathcal{L}) \\ &\xrightarrow{ex} H^{*+(p+r-n)}(N \times D^r, N \times D^r \setminus M; (p_N)^*\mathcal{L}) \\ &\xrightarrow{\cong} H^{*+(p+r-n)}(N \times D^r, N \times \partial D^r; (p_N)^*\mathcal{L}) \\ &\xrightarrow{(\bar{i}_M)^*} H^{*+(p+r-n)}(N \times D^r, N \times \partial D^r; (p_N)^*\mathcal{L}) \xrightarrow{((p_N)^*)^{-1}} H^{*+(p-n)}(N; \mathcal{L}), \end{aligned}$$

where $u \in H^{p+r-n}(U, U \setminus M; (p_U)^*\mathcal{O}_{f^*TN - TM})$ is the Thom class of the normal bundle ν , ex denotes the excision isomorphism, $p_N : N \times D^r \rightarrow N$ is the projection to the first factor, $\bar{i}_M : (N \times D^r, N \times \partial D^r) \rightarrow (N \times D^r, N \times D^r \setminus M)$ is the inclusion, and the final isomorphism comes from the Künneth theorem. It is known that the above definition does not depend on the choice of a particular embedding ε or r (for details, see [7], [8], [6], for example).

The following properties are known.

Lemma 2.1. (1) *If f is a proper map, then the twisted Gysin map $f_!$ coincides with the composition*

$$\begin{aligned} H^*(M; f^*\mathcal{L} \otimes \mathcal{O}_{f^*TN - TM}) &\xrightarrow{\smile [M]} H_{n-*}^c(M; f^*(\mathcal{L} \otimes \mathcal{O}_N)) \\ &\xrightarrow{f_*} H_{n-*}^c(N; \mathcal{L} \otimes \mathcal{O}_N) \xrightarrow{(\smile [N])^{-1}} H^{*+(p-n)}(N; \mathcal{L}), \end{aligned}$$

where H_*^c denotes the homology of closed support, and $[M] \in H_n^c(M; \mathcal{O}_M)$ and $[N] \in H_p^c(N; \mathcal{O}_N)$ are the fundamental classes of M and N respectively.

(2) *We have*

$$f_!(f^*x \smile y) = x \smile f_!(y)$$

for all $x \in H^*(N; \mathcal{L})$ and all $y \in H^*(M; f^*\mathcal{L}' \otimes \mathcal{O}_{f^*TN - TM})$, where \mathcal{L} and \mathcal{L}' are arbitrary local systems over N .

(3) *For smooth maps $f : M \rightarrow N$ and $g : N \rightarrow L$ with $\dim M = n$, $\dim N = p$ and $\dim L = \ell$, we have $(g \circ f)_! = g_! \circ f_! : H^*(M; (g \circ f)^*\mathcal{L} \otimes \mathcal{O}_{(g \circ f)^*TL - TM}) \rightarrow H^{*+(\ell-n)}(L; \mathcal{L})$ for any local system \mathcal{L} over L .*

Using the twisted Gysin map, we define the self-intersection class of a submanifold as follows.

Definition 2.2. Let Y be a smooth submanifold of a smooth manifold X of codimension κ . Let U be a tubular neighbourhood of Y in X with projection $p_U : U \rightarrow Y$, where U is identified with the total space of the normal bundle ν_Y of

Y in X . Consider the Thom class $u \in H^\kappa(U, U \setminus Y; (p_U)^* \mathcal{O}_{\nu_Y})$ of ν_Y . The twisted Euler class $e(\nu_Y) \in H^\kappa(Y; \mathcal{O}_{\nu_Y})$ of ν_Y is the image of u under the composition

$$H^\kappa(U, U \setminus Y; (p_U)^* \mathcal{O}_{\nu_Y}) \xrightarrow{(\bar{i}_Y)^*} H^\kappa(U; (p_U)^* \mathcal{O}_{\nu_Y}) \xrightarrow[\cong]{((p_U)^*)^{-1}} H^\kappa(Y; \mathcal{O}_{\nu_Y}),$$

where $\bar{i}_Y : U \rightarrow (U, U \setminus Y)$ is the inclusion. Then the self-intersection class $I(Y) \in H^{2\kappa}(X; \mathbf{Z})$ of Y is defined by $I(Y) = i_!(e(\nu_Y))$, where $i_! : H^\kappa(Y; \mathcal{O}_{\nu_Y}) \rightarrow H^{2\kappa}(X; \mathbf{Z})$ is the Gysin map induced by the inclusion $i : Y \rightarrow X$.

It is clear that when X is oriented, the homology class Poincaré dual to the self-intersection class $I(Y)$ is represented by the transverse intersection of Y and its small perturbation in X as a \mathbf{Z} -cycle (even if Y is non-orientable). Note that this integral cycle is well-defined as a homology class, which is denoted by $Y \cdot Y \in H_{n-2\kappa}(X; \mathbf{Z})$ with $n = \dim X$ (see [4, p. 583]). Note that $Y \cdot Y$ depends on the choice of an orientation for X , while $I(Y)$ does not.

3. PROOF OF THEOREM 1.2

In this section, we consider the self-intersection class of the singular set of a generic smooth map with corank one singularities and prove the formula (1.2).

Proof of Theorem 1.2. Let us first assume that $N = \mathbf{R}^p$. The general case will follow from this special case.

Let $\pi : G \rightarrow M$ be the Grassmannian bundle of unoriented $2k$ -planes in TM and ϕ the tautological $2k$ -plane bundle over G . Let us consider the vector bundle $\text{Hom}(\phi, \varepsilon^p)$ over G , where $\varepsilon^p = \pi^* f^* T\mathbf{R}^p$ is the trivial p -plane bundle. There is a natural section s of this vector bundle associated with f , which is defined by $s(x, H) = df_x|_H : H \rightarrow \mathbf{R}^p$ for $x \in M$ and $H \subset TM_x$. Our assumption on $j^1 f$ implies that s is transverse to the zero section (for details, see [23]). We set $\tilde{S}(f) = s^{-1}(0)$ and $\tilde{\pi} = \pi|_{\tilde{S}(f)}$. Furthermore, we denote by $j : \tilde{S}(f) \hookrightarrow G$ the inclusion. Note that $\tilde{\pi} : \tilde{S}(f) \rightarrow S(f)$ is a diffeomorphism and that the normal bundle ν_j of the embedding j is isomorphic to $j^* \text{Hom}(\phi, \varepsilon^p) \cong \text{Hom}(j^* \phi, j^* \varepsilon^p)$. Note also that $\mathcal{O}_{\text{Hom}(\phi, \varepsilon^p)} \cong (\mathcal{O}_\phi)^{\otimes p}$ and hence that $\mathcal{O}_{\nu_j} \cong j^*(\mathcal{O}_\phi)^{\otimes p}$, where for a local system \mathcal{L} , $\mathcal{L}^{\otimes p}$ denotes the p -fold tensor product of \mathcal{L} .

Let \tilde{U} be a tubular neighbourhood of $\tilde{S}(f)$ in G with projection $p_{\tilde{U}} : \tilde{U} \rightarrow \tilde{S}(f)$. Furthermore, let E denote the total space of the vector bundle $\text{Hom}(\phi, \varepsilon^p)$ and $\pi_E : E \rightarrow G$ the projection, where we consider G to be embedded in E as the zero section. Note that \tilde{U} can be identified with the total space of $j^* \text{Hom}(\phi, \varepsilon^p)$. Then, by considering the commutative diagram

$$\begin{array}{ccc} H^{2kp}(\tilde{U}, \tilde{U} \setminus \tilde{S}(f); (p_{\tilde{U}})^* \mathcal{O}_{\nu_j}) & \xrightarrow{ex} & H^{2kp}(G, G \setminus \tilde{S}(f); (\mathcal{O}_\phi)^{\otimes p}) \\ & \searrow (s|_{\tilde{U}})^* & \uparrow s^* \\ & & H^{2kp}(E, E \setminus G; (\pi_E)^* (\mathcal{O}_\phi)^{\otimes p}) \\ & \searrow (\bar{i}_{\tilde{S}(f)})^* & \\ & & H^{2kp}(G; (\mathcal{O}_\phi)^{\otimes p}) \\ & & \cong \uparrow s^* = ((\pi_E)^*)^{-1} \\ & \searrow (\bar{i}_G)^* & \\ & & H^{2kp}(E; (\pi_E)^* (\mathcal{O}_\phi)^{\otimes p}), \end{array}$$

we see that

$$(3.1) \quad j_!(1) = e(\text{Hom}(\phi, \varepsilon^p)) = e(\phi)^p \in H^{2kp}(G; (\mathcal{O}_\phi)^{\otimes p}),$$

where $\bar{i}_{\tilde{S}(f)} : G \rightarrow (G, G \setminus \tilde{S}(f))$ and $\bar{i}_G : E \rightarrow (E, E \setminus G)$ are the inclusion maps.

Let $K = \ker df$ and $Q = \operatorname{coker} df$ be the kernel bundle of rank $2k$ and the cokernel bundle of rank 1 defined over $S(f)$, respectively. Note that $\operatorname{Hom}(K, Q)$ is isomorphic to the normal bundle ν of $i : S(f) \hookrightarrow M$ (for example, see [5], [15]). Over $\tilde{S}(f)$, we have $\tilde{\pi}^*\nu \cong \operatorname{Hom}(\tilde{\pi}^*K, \tilde{\pi}^*Q)$ and $\tilde{\pi}^*K \cong j^*\phi$. Hence, we have $\tilde{\pi}^*\mathcal{O}_\nu \cong \tilde{\pi}^*\mathcal{O}_K \cong j^*\mathcal{O}_\phi$, since the rank of K is even.

Lemma 3.1. *We have*

$$(3.2) \qquad e(\tilde{\pi}^*\nu) \equiv e(j^*\phi) \in H^{2k}(\tilde{S}(f); j^*\mathcal{O}_\phi) \quad (\text{modulo } 2\text{-torsion}).$$

Proof. When $\tilde{\pi}^*Q$ is trivial, (3.2) obviously holds even without taking modulo 2-torsion, since $\tilde{\pi}^*\nu \cong \tilde{\pi}^*K \cong j^*\phi$. When $\tilde{\pi}^*Q$ is not trivial, let $\widehat{\pi} : \widehat{S}(f) \rightarrow \tilde{S}(f)$ be the double covering corresponding to $w_1(\tilde{\pi}^*Q) \in H^1(\tilde{S}(f); \mathbf{Z}_2)$, the first Stiefel-Whitney class. Then, since $\widehat{\pi}^*\tilde{\pi}^*Q$ is trivial, we have $e(\widehat{\pi}^*\tilde{\pi}^*\nu) = e(\widehat{\pi}^*j^*\phi) \in H^{2k}(\widehat{S}(f); \widehat{\pi}^*j^*\mathcal{O}_\phi)$. Then by Lemma 2.1 (2), we have

$$2e(\tilde{\pi}^*\nu) = \widehat{\pi}_!(e(\widehat{\pi}^*\tilde{\pi}^*\nu)) = \widehat{\pi}_!(e(\widehat{\pi}^*j^*\phi)) = 2e(j^*\phi),$$

since $\widehat{\pi}_!(1) = 2$. Thus (3.2) holds. □

Lemma 3.2. *We have*

$$(\mathcal{O}_\phi)^{\otimes n-2k} \otimes \mathcal{O}_G \cong \pi^*\mathcal{O}_M.$$

Proof. Recall that the Grassmann manifold $G_{2k}(\mathbf{R}^n)$ consisting of $2k$ -planes in \mathbf{R}^n is orientable if and only if $n - 2k$ is even. Furthermore, its orientation remains invariant under the orientation reversal of \mathbf{R}^n . Hence, when $n - 2k$ is even, we have

$$(\mathcal{O}_\phi)^{\otimes n-2k} \otimes \mathcal{O}_G \cong \mathcal{O}_G \cong \pi^*\mathcal{O}_M.$$

When $n - 2k$ is odd, let γ be the tautological $2k$ -plane bundle over $G_{2k}(\mathbf{R}^n)$. Then the orientation local system \mathcal{O}_γ is isomorphic to $\mathcal{O}_{G_{2k}(\mathbf{R}^n)}$, since $TG_{2k}(\mathbf{R}^n)$ is isomorphic to $\operatorname{Hom}(\gamma, \gamma^\perp)$, where γ^\perp is the orthogonal complement of γ . Hence, we have $(\mathcal{O}_\phi)^{\otimes n-2k} \otimes \mathcal{O}_G \cong \mathcal{O}_\phi \otimes \mathcal{O}_G \cong \pi^*\mathcal{O}_M$. □

By the above lemma, $\pi : G \rightarrow M$ induces the Gysin map

$$\pi_! : H^*(G; (\mathcal{O}_\phi)^{\otimes n-2k}) \cong H^*(G; \mathcal{O}_G \otimes \pi^*\mathcal{O}_M) \rightarrow H^{*-2k(n-2k)}(M; \mathbf{Z}).$$

Lemma 3.3. *We have*

$$\pi_!(e(\phi)^{n-2k}) = 1,$$

where $\pi_! : H^{2k(n-2k)}(G; (\mathcal{O}_\phi)^{\otimes n-2k}) \rightarrow H^0(M; \mathbf{Z})$ is the Gysin map induced by $\pi : G \rightarrow M$.

Proof. The proof is similar to that in Porteous' paper [23, Proposition 0.3], as follows.

We have only to consider the equality over a fibre $G(TM_x) = \pi^{-1}(x)$, that is, over the Grassmannian of TM_x , since the diagram

$$\begin{array}{ccc} H^{2k(n-2k)}(G; (\mathcal{O}_\phi)^{\otimes n-2k}) & \xrightarrow{\pi_!} & H^0(M; \mathbf{Z}) \\ (i_{G(TM_x)})^* \downarrow & & \downarrow (i_x)^* \\ H^{2k(n-2k)}(G(TM_x); (\mathcal{O}_{\phi(x)})^{\otimes n-2k}) & \xrightarrow{(\pi_x)_!} & H^0(x; \mathbf{Z}) \end{array}$$

is commutative, where $i_{G(TM_x)} : G(TM_x) \rightarrow G$ and $i_x : x \rightarrow M$ are the inclusions, $\pi_x = \pi|_{G(TM_x)} : G(TM_x) \rightarrow x$, and $\phi(x)$ is the restriction of ϕ to $G(TM_x)$.

Fix a nonzero vector $v \in TM_x$. Then it induces a section s_v of $\phi(x)$ defined by

$$s_v(H) = [v] \in TM_x/H^\perp \equiv H \quad \text{for } H \in G(TM_x),$$

where H^\perp is the orthogonal complement of H with respect to a fixed inner product on TM_x . The zero set $s_v^{-1}(0) (= \{H : v \in H^\perp\})$ represents the homology class in $H_{2k(n-2k)-2k}(G(TM_x); \mathcal{O}_{G(TM_x)} \otimes \mathcal{O}_{\phi(x)})$ Poincaré dual to the Euler class $e(\phi(x))$. If we take a set of linearly independent $(n-2k)$ -vectors of TM_x , then we get $n-2k$ generically linearly independent sections $s_i, i = 1, 2, \dots, n-2k$. The intersection of all $s_i^{-1}(0)$ consists of a unique point, which corresponds to the $2k$ -plane orthogonal to the space spanned by the $(n-2k)$ -vectors. Thus we have that the homology class in $H_0(G(TM_x); \mathcal{O}_{G(TM_x)} \otimes (\mathcal{O}_{\phi(x)})^{\otimes n-2k}) \cong H_0(G(TM_x); \mathbf{Z})$ Poincaré dual to $e(\phi(x))^{n-2k}$ is represented by a point. Hence, the result follows by Lemma 2.1 (1). \square

Let us go back to the proof of Theorem 1.2. We fix a Riemannian metric on M and consider the orthogonal decomposition $\pi^*TM = \phi \oplus \phi^\perp$. Note that $e(\phi)^2 = p_k(\phi)$ (for example, see [21, §15]). By using the product formula

$$p(\pi^*TM) \equiv p(\phi)p(\phi^\perp) \quad (\text{modulo 2-torsion})$$

for Pontrjagin classes together with Lemma 2.1 and (3.1), we have

$$\begin{aligned} j_!(e(j^*\phi)) &= e(\phi)j_!(1) = e(\phi)^{p+1} = p_k(\phi)e(\phi)^{n-2k} \\ &\equiv \left(p_k(\pi^*TM) - \sum_{\ell=0}^{k-1} p_\ell(\phi)p_{k-\ell}(\phi^\perp) \right) e(\phi)^{n-2k} \\ &\equiv p_k(\pi^*TM)e(\phi)^{n-2k} - \left(\sum_{\ell=0}^{k-1} p_\ell(\phi)(p_{k-\ell}(\phi^\perp)p_k(\phi)) \right) e(\phi)^{n-2k-2} \\ &\equiv p_k(\pi^*TM)e(\phi)^{n-2k} - \left(\sum_{\ell_1=0}^{k-1} p_{\ell_1}(\phi) \right. \\ &\quad \left. \left(p_{2k-\ell_1}(\pi^*TM) - \sum_{\ell_2=0}^{k-1} p_{\ell_2}(\phi)p_{2k-\ell_1-\ell_2}(\phi^\perp) \right) \right) e(\phi)^{n-2k-2} \\ &\equiv \dots \\ &\equiv p_k(\pi^*TM)e(\phi)^{n-2k} + T_1 + T_2 \quad (\text{modulo 2-torsion}). \end{aligned}$$

Here we set

$$(3.3) \quad T_1 = (-1)^s \sum_{I_s} p_{j_s}(\pi^*TM) p_{I_s}(\phi) e(\phi)^{n-2k-2s},$$

$$(3.4) \quad T_2 = (-1)^{n_0+1} \sum_{I_{n_0+1}} p_{I_{n_0+1}}(\phi) p_{k_0}(\phi^\perp) e(\phi)^{n-2k-2n_0},$$

where $n_0 = [(n-2k)/2]$ is the greatest integer not exceeding $(n-2k)/2$, the sum in (3.3) runs over all multi-indices

$$I_s = (\ell_1, \dots, \ell_s) \quad \text{with} \quad 0 \leq \ell_1, \dots, \ell_s \leq k-1 \quad (1 \leq s \leq [(n-2k)/2] = n_0),$$

the sum in (3.4) runs over all multi-indices

$$I_{n_0+1} = (\ell_1, \dots, \ell_{n_0+1}) \quad \text{with} \quad 0 \leq \ell_1, \dots, \ell_{n_0+1} \leq k-1,$$

$$p_{I_t} = p_{\ell_1} p_{\ell_2} \cdots p_{\ell_t}, \quad j_s = (s+1)k - (\ell_1 + \cdots + \ell_s), \quad \text{and} \quad k_0 = (n_0+1)k - (\ell_1 + \cdots + \ell_{n_0+1}).$$

Since $\ell_t \leq k-1$ for all $1 \leq t \leq n_0+1$, we have $k_0 \geq n_0+1 > (n-2k)/2$. Therefore, $p_{k_0}(\phi^\perp)$ and hence T_2 vanishes. Furthermore, the degree of $p_{I_s}(\phi)e(\phi)^{n-2k-2s}$ is strictly smaller than $2k(n-2k)$. Therefore, its image by the Gysin map $\pi_!$ vanishes,

and hence we have $\pi_1 T_1 = 0$ by Lemma 2.1 (2). Thus, by Lemmas 3.3 and 2.1, we have

$$\pi_1 j_! (e(j^* \phi)) \equiv p_k(TM) \pi_! (e(\phi)^{n-2k}) = p_k(TM) \pmod{2\text{-torsion}}.$$

On the other hand, by (3.2) and Lemma 2.1 we have

$$\begin{aligned} \pi_1 j_! (e(j^* \phi)) &= i_! \widetilde{\pi}_! (e(j^* \phi)) \equiv i_! \widetilde{\pi}_! (e(\widetilde{\pi}^* \nu)) \pmod{2\text{-torsion}} \\ &= i_! (e(\nu) \widetilde{\pi}_! (1)) = i_! (e(\nu)). \end{aligned}$$

Thus, by definition of the self-intersection class, we get

$$I(S(f)) = i_! (e(\nu)) \equiv p_k(TM) \pmod{2\text{-torsion}}.$$

This completes the proof for the case $N = \mathbf{R}^p$.

In the case of general N , we can show the assertion similarly in a standard way, which is sketched as follows (for more details, see [12] for example). Take a vector bundle ξ over M such that $f^*TN \oplus \xi$ is trivial (for instance, taking an embedding of N into a Euclidean space of sufficiently high dimension, set ξ to be the pull-back of the normal bundle of the embedding). Note that ξ defines the element $-f^*TN$ in the K -group. In the above proof, TM and df (i.e., $j^1 f$) should be replaced by $TM \oplus \xi$ and $df \oplus \text{id}_\xi : TM \oplus \xi \rightarrow f^*TN \oplus \xi$, respectively. Finally, we get

$$i_! (e(\nu)) \equiv p_k(TM \oplus \xi) \equiv p_k(TM - f^*TN) \pmod{2\text{-torsion}}.$$

This completes the proof of Theorem 1.2. □

Remark 3.4. When the normal bundle ν of the embedding $i : S(f) \hookrightarrow M$ is orientable, the Euler class $e(\nu) \in H^{2k}(S(f); \mathbf{Z})$ is a special case of the incident class defined by Rimányi [24], i.e., $e(\nu) = I(\Sigma^{2k}, \Sigma^{2k})(f)$. Note that our definition of the twisted Euler class uses the orientation local system so that everything works even if ν is non-orientable.

Now the formula (1.3) follows directly from Theorem 1.2, since $\langle I(S(f)), [M] \rangle = S(f) \cdot S(f)$ by the definition of $I(S(f))$ and every orientable 3-manifold is parallelizable, where $[M] \in H_4(M; \mathbf{Z})$ is the fundamental class of M , and $\langle \cdot, \cdot \rangle$ is the Kronecker product.

Remark 3.5. When $n - p + 1$ is odd, $e(\nu) \in H^{n-p+1}(S(f); \mathcal{O}_\nu)$ is of order at most two, since the dimension of a fibre of ν is odd. Hence $I(S(f)) = i_! (e(\nu))$ is an element of order at most two or, in other words, it vanishes modulo 2-torsion.

Remark 3.6. Fehér and Rimányi in [13] observed that if a singularity set $\eta(f)$ of type η is an orientable closed submanifold of an orientable manifold, then the self-intersection class of $\eta(f)$ is equal to the Thom polynomial of the complexified singularity $\eta_{\mathbf{C}}(f_{\mathbf{C}})$ multiplied by $(-1)^{m(m-1)/2}$, where m is the (real) codimension of η . It seems possible to generalize this to the non-orientable case, as we have done here for the singularity type A_1 . In fact, for complex analytic maps $M^n \rightarrow N^{n-2k+1}$, the Thom polynomial for an A_1 -type singularity coincides with $c_{2k}(TM - f^*TN)$. We conjecture that our formula would be true without taking modulo 2-torsion, and also for the case $\Sigma^\ell(f) \neq \emptyset$ for some $\ell \geq 2k + 1$.

Remark 3.7. For a generic map $f : M \rightarrow N$ of a closed n -dimensional manifold into a p -dimensional manifold with $n - p + 1 = 2k$, $k \geq 1$, the Thom polynomial

$\text{Tp}(A_1(f))$ coincides with the $2k$ -th Stiefel-Whitney class $w_{2k}(TM - f^*TN) \in H^{2k}(M; \mathbf{Z}_2)$ by Thom [31]. Hence, we have

$$I(S(f)) \equiv w_{2k}(TM - f^*TN)^2 \pmod{2}.$$

On the other hand, we have the basic congruence

$$w_{2k}^2 \equiv p_k \pmod{2}$$

(see [21, Problem 15-A], for example). This means that the difference $I(S(f)) - p_k(TM - f^*TN)$ is a multiple of two.

Remark 3.8. In Theorem 1.2, we have assumed that the source manifold M is compact. However, this assumption is not necessary, as long as the map $f : M \rightarrow N$ is proper and we use the homology of closed support instead of the usual homology.

4. NONEXISTENCE OF FOLD MAPS

In this section, we give some necessary conditions for the existence of fold maps as an application of our formula (1.2). Let us first recall the following.

Definition 4.1. For a smooth map $f : M \rightarrow N$ with $n = \dim M \geq \dim N = p$, a point $q \in M$ is a *fold singularity* (or an A_1 -type singularity) of f if f is of the form

$$(4.1) \quad (x_1, \dots, x_n) \mapsto (x_1, \dots, x_{p-1}, \pm x_p^2 \pm \dots \pm x_n^2)$$

with respect to appropriate coordinates around q and $f(q)$. Let λ' be the number of negative signs appearing in the p -th component of the right-hand side of (4.1). Then, $\max\{\lambda', n-p+1-\lambda'\}$ is called the *reduced index* of q , which does not depend on a particular choice of coordinates (for details, see [20]). If the singular set $S(f)$ of f is empty or consists only of fold singularities, f is called a *fold map*.¹

Now let us prove Theorem 1.3.

Proof of Theorem 1.3. Suppose that there is a fold map $f : M \rightarrow N$ homotopic to g . For $\lambda = k, k+1, \dots, n-p+1$, let $S_\lambda(f)$ be the set of fold singularities of f of reduced index λ . Note that $S(f)$ is the disjoint union of $S_k(f), S_{k+1}(f), \dots, S_{n-p+1}(f)$ and that each $S_\lambda(f)$ consists of some connected components of $S(f)$. As is easily observed, each $S_\lambda(f)$ is a $(p-1)$ -dimensional regular submanifold of M , and $f|_{S_\lambda(f)}$ is an immersion for each λ . Furthermore, the normal bundle of the immersion is trivial for $\lambda \neq k$ (for example, see [25]). This implies that $\mathcal{O}_{S_\lambda(f)} \cong (i_\lambda)^* f^* \mathcal{O}_N$ for $\lambda \neq k$, where $i_\lambda : S_\lambda(f) \rightarrow M$ is the inclusion. Let $x_\lambda \in H^{2k}(M; \mathcal{O}_{f^*TN-TM}) = H^{2k}(M; \mathcal{O}_{g^*TN-TM})$ be the cohomology class Poincaré dual to the homology class in $H_{p-1}(M; f^* \mathcal{O}_N)$ represented by $S_\lambda(f)$, $\lambda \neq k$. Then, the self-intersection homology class $S_\lambda(f) \cdot S_\lambda(f) \in H_{n-4k}(M; \mathcal{O}_M)$ and the cohomology class $x_\lambda \smile x_\lambda \in H^{4k}(M; \mathbf{Z})$ are Poincaré dual to each other, and the self-intersection class $I(S_\lambda(f))$ coincides with $x_\lambda \smile x_\lambda$.

For $\lambda = k$, let ν_k be the normal bundle of $S_k(f)$ in M . By [25], its structure group can be reduced to the semi-direct product $G = (O(k) \times O(k)) \rtimes \mathbf{Z}_2 \subset O(2k)$. Since $O(k) \times O(k)$ is a subgroup of G of index two, the structure group of the pull-back $\pi_k^* \nu_k$ of ν_k by an appropriate double covering $\pi_k : \tilde{S}_k(f) \rightarrow S_k(f)$ can be reduced to $O(k) \times O(k)$. This implies that $\pi_k^* \nu_k$ splits into the Whitney sum $\xi_1 \oplus \xi_2$ for some k -plane bundles ξ_1 and ξ_2 over $\tilde{S}_k(f)$.

¹In [15], such a map is called a *submersion with folds*.

Since k is odd by our assumption, the twisted Euler classes of ξ_1 and ξ_2 are elements of order at most two, and hence so is the Euler class $e(\pi_k^*\nu_k)$ of $\pi_k^*\nu_k$. Since $\pi_{k!}(e(\pi_k^*\nu_k)) = e(\nu_k)\pi_{k!}(1) = 2e(\nu_k)$, we see that $4e(\nu_k)$ vanishes. This implies that the self-intersection class $I(S_k(f))$ vanishes modulo 4-torsion.

Since $x_\lambda \smile x_{\lambda'} = 0$ for $\lambda \neq \lambda'$ and

$$I(S(f)) = \sum_{\lambda=k}^{n-p+1} I(S_\lambda(f)),$$

we conclude that the cohomology class

$$x = \sum_{\lambda=k+1}^{n-p+1} x_\lambda \in H^{2k}(M; \mathcal{O}_{g^*TN-TM})$$

satisfies

$$x \smile x \equiv I(S(f)) \pmod{4\text{-torsion}}.$$

Hence, by Theorem 1.2, the result follows. This completes the proof of Theorem 1.3. \square

As an important corollary, we have the following, which shows the effectiveness of the formula (1.3).

Corollary 4.2. *Let M be a closed oriented 4-manifold whose intersection form is isomorphic either to $\pm\langle 1 \rangle$ or to $\pm(\langle 1 \rangle \oplus \langle 1 \rangle)$. Then, there exists no fold map $f : M \rightarrow N$ for any orientable 3-manifold N . In other words, every generic map $f : M \rightarrow N$ necessarily has cusp singularities.*

Proof. Suppose that there is a fold map $f : M \rightarrow N$. Then by Theorem 1.3, there exists an element $x \in H^2(M; \mathbf{Z})$ such that

$$x \smile x = p_1(TM - f^*TN) = p_1(M),$$

since every orientable 3-manifold is parallelizable. When the intersection form of M is isomorphic to $\pm\langle 1 \rangle$, this implies that there exists an integer ℓ such that $\ell^2 = 3$, since $p_1[M] = 3\sigma(M)$. This is a contradiction. Similarly, when the intersection form of M is isomorphic to $\pm(\langle 1 \rangle \oplus \langle 1 \rangle)$, there must exist integers ℓ_1 and ℓ_2 such that $\ell_1^2 + \ell_2^2 = 6$, which is a contradiction again. Hence, there exists no fold map $f : M \rightarrow N$. \square

Remark 4.3. In [27], the second author obtained the special case of Theorem 1.3 for $(n, p) = (4, 3)$ by using a different method. Corollary 4.2 was also obtained there. (In fact, in [27], it was proved that the sufficient condition for the nonexistence of fold maps mentioned in Corollary 4.2 is also necessary.) Note that Corollary 4.2 generalizes Theorem 1.1 and the main theorem of [30]. Note also that Akhmetiev and Sadykov [1] recently gave another proof of Theorem 1.1 from a slightly different point of view.

By using Theorem 1.3, we also have the following result for general dimensions.

Corollary 4.4. *For every dimension pair (n, p) such that $n - p + 1 = 2k$ for a positive odd integer k with $4k \leq n$, there exists a closed connected orientable manifold of dimension n that admits no fold map into any p -dimensional manifold N such that $p_i(N) = 0$ for all $1 \leq i \leq k$.*

Proof. Set

$$M = \begin{cases} \mathbf{CP}^{2k} \times S^{n-4k}, & n - 4k \geq 2, \\ \mathbf{CP}^{2k-1} \times S^{n-4k+2}, & n - 4k = 1, \\ \mathbf{CP}^{2k}, & n = 4k. \end{cases}$$

Note that its cohomology ring satisfies

$$H^*(M; \mathbf{Z}) \cong \begin{cases} \mathbf{Z}[a, b]/(a^{2k+1}, b^2), & n - 4k \geq 2, \\ \mathbf{Z}[a, b]/(a^{2k}, b^2), & n - 4k = 1, \\ \mathbf{Z}[a]/(a^{2k+1}), & n = 4k, \end{cases}$$

where a corresponds to a generator of $H^2(\mathbf{CP}^{2k}; \mathbf{Z}) \cong \mathbf{Z}$ for $n - 4k \neq 1$ (or a generator of $H^2(\mathbf{CP}^{2k-1}; \mathbf{Z}) \cong \mathbf{Z}$ for $n - 4k = 1$), and b corresponds to a generator of $H^{n-4k}(S^{n-4k}; \mathbf{Z}) \cong \mathbf{Z}$ for $n - 4k \geq 2$ (or a generator of $H^{n-4k+2}(S^{n-4k+2}; \mathbf{Z}) \cong \mathbf{Z}$ for $n - 4k = 1$). Note also that the total Pontrjagin class $p(M)$ of M satisfies

$$p(M) = \begin{cases} (1 + a^2)^{2k+1}, & n - 4k \neq 1, \\ (1 + a^2)^{2k}, & n - 4k = 1 \end{cases}$$

(for example, see [21]).

Suppose that there exists a fold map $f : M \rightarrow N$ for some p -dimensional manifold N such that $p_i(N) = 0$ for $1 \leq i \leq k$. Since M is simply connected, any local system over M is trivial. Then by Theorem 1.3, there exists an element $x \in H^{2k}(M; \mathbf{Z})$ such that

$$x \smile x = p_k(TM - f^*TN) = p_k(M) = \begin{cases} \binom{2k+1}{k} a^{2k}, & n - 4k \neq 1, \\ \binom{2k}{k} a^{2k}, & n - 4k = 1, \end{cases}$$

since $H^*(M; \mathbf{Z})$ is torsion free. This implies that the integer

$$\binom{2k+1}{k} = \frac{(2k+1)!}{(k+1)!k!} \quad \text{or} \quad \binom{2k}{k} = \frac{(2k)!}{k!k!}$$

must be a square, which is a contradiction by a result of Erdős [11] (see also [16]). Hence, M is a desired manifold. This completes the proof. \square

Remark 4.5. In the above corollary, if n is even and $k \geq 3$, then we can prove that $\mathbf{CP}^{n/2}$ is also a desired manifold except possibly for $(n, p) = (98, 93)$, since the binomial coefficient

$$\binom{n/2+1}{k}$$

is a square if and only if $n = 98$ and $k = 3$ (see [16]).

Corollary 4.6. *Let (n, p) be a dimension pair such that $n - p + 1 = 2k$ for a positive odd integer k with $4k < n$. Then, for every closed oriented n -dimensional manifold M , there exists a closed connected oriented n -dimensional manifold M' oriented cobordant to M such that M' admits no fold map into any p -dimensional manifold N with $p_i(N) = 0$ for all $1 \leq i \leq k$.*

Proof. Let M_0 be the n -dimensional manifold given by Corollary 4.4. Furthermore, let M_1 be a connected and simply connected manifold oriented cobordant to M and set $M' = M_1 \# M_0 \# \overline{M_0}$, where $\overline{M_0}$ denotes the manifold M_0 with orientation reversed.

Then, since we have $n > 4k$ by our assumption, by an argument similar to that in the proof of the above corollary, we see that there exists no element $x \in H^{2k}(M'; \mathbf{Z})$ such that $x \smile x \equiv p_k(M')$ modulo 4-torsion. Hence, by Theorem 1.3, M' admits no fold map into N . Since M' is oriented cobordant to M , the result follows. \square

Remark 4.7. Surprisingly enough, for $(n, p) = (4, 3)$, the conclusion of the above corollary does not hold in general. In fact, if a closed oriented 4-manifold M has signature $\sigma(M) \neq \pm 1, \pm 2$, then M always admits a fold map into \mathbf{R}^3 (see [27]). Note that, in this case, $k = 1$ and $4k = n$.

Remark 4.8. In [19], it has been shown that if M is a closed manifold of odd Euler characteristic, then M cannot admit any fold map into \mathbf{R}^p for $p \neq 1, 3, 7$ (see also [28]). We can use this to obtain a result similar to Corollary 4.4 for other dimension pairs as well. However, such a result is not useful for the proof of Corollary 4.6.

Example 4.9. Let us consider the 4-dimensional complex projective space \mathbf{CP}^4 . By [19], if \mathbf{CP}^4 admits a fold map into \mathbf{R}^p , then p must be equal to 1, 3 or 7, since \mathbf{CP}^4 has odd Euler characteristic. Clearly, it admits a fold map into \mathbf{R} . However, it cannot admit a fold map into \mathbf{R}^7 , since

$$\binom{5}{1} = 5$$

is not a square (for details, see the proof of Corollary 4.4 or Remark 4.5). We do not know if \mathbf{CP}^4 admits a fold map into \mathbf{R}^3 or not.

Similar observations hold also for \mathbf{CP}^6 , $\mathbf{CP}^2 \times \mathbf{CP}^2$, \mathbf{HP}^2 , etc. (refer to [21] for the description of their Pontrjagin classes). Details are left to the reader.

Remark 4.10. Let $f : M \rightarrow N$ be a fold map of a closed orientable n -dimensional manifold into an orientable p -dimensional manifold such that $n - p + 1 = 2k$. Let us suppose that the singular set $S(f)$ of f is orientable. We give an arbitrary orientation to $S(f)$ and let $v \in H^{2k}(M; \mathbf{Z})$ be the cohomology class Poincaré dual to the homology class represented by $S(f)$. Then by [31], the modulo two reduction of v coincides with the Stiefel-Whitney class $w_{2k}(TM - f^*TN) \in H^{2k}(M; \mathbf{Z}_2)$. Furthermore, by Theorem 1.2, $v \smile v \equiv p_k(TM - f^*TN)$ modulo 2-torsion. Summarizing, we see that there exists an element $v \in H^{2k}(M; \mathbf{Z})$ whose modulo two reduction coincides with $w_{2k}(TM - f^*TN)$ such that $v \smile v \equiv p_k(TM - f^*TN)$ modulo 2-torsion.

The above result can also be proved as follows. By Ando [2], there exists a fold map $f : M \rightarrow N$ homotopic to a given smooth map $g : M \rightarrow N$ such that $S(f)$ is orientable if and only if there exists a fibrewise epimorphism $\varphi : TM \oplus \varepsilon^1 \rightarrow g^*TN$, where ε^1 is the trivial line bundle over M . Set $\eta = \ker \varphi$, which is an orientable vector bundle of rank $n - p + 1 = 2k$, and let v denote the Euler class of η . Then its modulo two reduction coincides with $w_{2k}(TM - f^*TN)$, and $v \smile v = p_k(TM - f^*TN)$ (for details, see [21], for example).

The above observation suggests that the formula (1.2) should be true without taking modulo 2-torsion.

Remark 4.11. So far, we have obtained several nonexistence results for fold maps. For the existence results, refer to the works of Levine [20], Èliášberg [9], [10] and Ando [2]. For the dimension pair $(n, p) = (4, 3)$, see [27].

REFERENCES

- [1] P. M. Akhmetiev and R. R. Sadykov, *A remark on elimination of singularities for mappings of 4-manifolds into 3-manifolds*, preprint, 2002.
- [2] Y. Ando, *Existence theorems of fold-maps*, preprint, 2001.
- [3] V. I. Arnol'd, V. V. Goryunov, O. V. Lyashko, and V. A. Vasil'ev, *Singularity Theory I*, Encyclopaedia of Mathematical Sciences, vol. 6, Dynamical Systems VI, Springer-Verlag, Berlin, 1993. MR **94b**:58018
- [4] M. F. Atiyah and I. M. Singer, *The index of elliptic operators: III*, Ann. of Math. **87** (1968), 546–604. MR **38**:5245
- [5] J. Boardman, *Singularities of differentiable maps*, Inst. Hautes Etudes Sci. Publ. Math. **33** (1967), 21–57. MR **37**:6945
- [6] G. E. Bredon, *Sheaf theory*, second edition, Graduate Texts in Math. 170, Springer-Verlag, New York, 1997. MR **98g**:55005
- [7] A. Dold, *Lectures on algebraic topology*, Springer-Verlag, Heidelberg, 1972. MR **54**:3685
- [8] E. Dyer, *Cohomology theories*, W. A. Benjamin, Inc., New York, 1969. MR **42**:8780
- [9] J. M. Éliašberg, *On singularities of folding type*, Math. USSR-Izv. **4** (1970), 1119–1134. MR **43**:4051
- [10] J. M. Éliašberg, *Surgery of singularities of smooth mappings*, Math. USSR-Izv. **6** (1972), 1302–1326. MR **49**:4021
- [11] P. Erdős, *On a Diophantine equation*, J. London Math. Soc. **26** (1951), 176–178. MR **12**:804d
- [12] L. Fehér and R. Rimányi, *Calculation on Thom polynomials for group actions*, preprint, 2000, arXiv:math.AG/0009085.
- [13] L. Fehér and R. Rimányi, *Thom polynomials with integer coefficients*, to appear in Illinois J. Math.
- [14] T. Fukuda, *Topology of folds, cusps and Morin singularities*, in “A Fete of Topology”, eds. Y. Matsumoto, T. Mizutani and S. Morita, Academic Press, 1987, pp. 331–353. MR **89a**:58015
- [15] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Graduate Texts in Math. 14, Springer-Verlag, New York, 1973. MR **49**:6269
- [16] K. Györy, *On the Diophantine equation $\binom{n}{k} = x^l$* , Acta Arith. **80** (1997), 289–295. MR **98f**:11028
- [17] A. Haefliger et A. Kosinski, *Un théorème de Thom sur les singularités des applications différentiables*, Séminaire H. Cartan, Ecole Norm. Sup., 1956/57, Exposé no. 8. MR **23**:A1382b
- [18] M. Hirsch, *Immersions of manifolds*, Trans. Amer. Math. Soc. **93** (1959), 242–276. MR **22**:9980
- [19] S. Kikuchi and O. Saeki, *Remarks on the topology of folds*, Proc. Amer. Math. Soc. **123** (1995), 905–908. MR **95j**:57032
- [20] H. Levine, *Elimination of cusps*, Topology **3** (suppl. 2) (1965), 263–296. MR **31**:756
- [21] J. Milnor and J. Stasheff, *Characteristic classes*, Ann. of Math. Studies, vol. 76, Princeton Univ. Press, Princeton, 1974. MR **55**:13428
- [22] A. Phillips, *Submersions of open manifolds*, Topology **6** (1967), 171–206. MR **34**:8420
- [23] I. R. Porteous, *Simple singularities of maps*, Proceedings of Liverpool Singularities I, Lecture Notes in Math., vol. 192, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1971, pp. 286–307. MR **45**:2723
- [24] R. Rimányi, *Thom polynomials, symmetries and incidences of singularities*, Invent. Math. **143** (2001), 499–521. MR **2001k**:58082
- [25] O. Saeki, *Notes on the topology of folds*, J. Math. Soc. Japan **44** (1992), 551–566. MR **93f**:57037
- [26] O. Saeki, *Studying the topology of Morin singularities from a global viewpoint*, Math. Proc. Cambridge Philos. Soc. **117** (1995), 223–235. MR **96g**:57030
- [27] O. Saeki, *Fold maps on 4-manifolds*, preprint.
- [28] O. Saeki and K. Sakuma, *Maps with only Morin singularities and the Hopf invariant one problem*, Math. Proc. Cambridge Philos. Soc. **124** (1998), 501–511. MR **99h**:57058
- [29] K. Sakuma, *On special generic maps of simply connected $2n$ -manifolds into \mathbf{R}^3* , Topology Appl. **50** (1993), 249–261. MR **94d**:57056

- [30] K. Sakuma, *A note on nonremovable cusp singularities*, Hiroshima Math. J. **31** (2001), 461–465. MR **2002h:57037**
- [31] R. Thom, *Les singularités des applications différentiables*, Ann. Inst. Fourier (Grenoble) **6** (1955–56), 43–87. MR **19:310a**

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, FACULTY OF SCIENCE, KAGOSHIMA UNIVERSITY, KOORIMOTO, KAGOSHIMA 890-0065, JAPAN

E-mail address: ohmoto@sci.kagoshima-u.ac.jp

FACULTY OF MATHEMATICS, KYUSHU UNIVERSITY, HAKOZAKI, FUKUOKA 812-8581, JAPAN

E-mail address: saeki@math.kyushu-u.ac.jp

DEPARTMENT OF MATHEMATICS AND PHYSICS, FACULTY OF SCIENCE AND TECHNOLOGY, KINKI UNIVERSITY, HIGASHI-OSAKA, OSAKA 577-8502, JAPAN

E-mail address: sakuma@math.kindai.ac.jp

ERRATUM TO “ARENS REGULARITY OF THE ALGEBRA $A \hat{\otimes} B$ ”

A. ÜLGER

Professors Edmond Granirer and Graham Colin have informed me that Theorem 4.20 in the last section of [1], which states that the projective tensor product $C(K) \hat{\otimes} C(S)$ of two commutative C^* -algebras is Arens regular, is not correct. The reason, as they have rightly told me, is that, if G is a compact group, then the algebra $C(G) \hat{\otimes} C(G)$ contains an isomorphic copy of the Fourier algebra $A(G)$, which is known to be not Arens regular. Consequently, $C(G) \hat{\otimes} C(G)$ is not Arens regular either. Unfortunately, this mistake also affects the subsequent results (from 4.21 to 4.27) of the last section. By this erratum I acknowledge this mistake and let it be known to the people working in this field.

REFERENCES

- [1] A. Ülger, *Arens regularity of the algebra $A \hat{\otimes} B$* , Trans. Amer. Math. Soc. **305** (1988), 623–639.
 MR **89c**:46064

DEPARTMENT OF MATHEMATICS, KOÇ UNIVERSITY, 34450, SARIYER, ISTANBUL
E-mail address: `aulger@ku.edu.tr`

Received by the editors December 12, 2002.

1991 *Mathematics Subject Classification*. Primary 46H05; Secondary 46J15, 46M05.

ERRATUM TO “SPHERICAL CLASSES AND THE ALGEBRAIC TRANSFER”

NGUYỄN H. V. HUNG

Let Q_0S^0 be the basepoint component of $QS^0 = \lim_n \Omega^n S^n$. A spherical class in Q_0S^0 is an element belonging to the image of the Hurewicz homomorphism:

$$H : \pi_*^s(S^0) \cong \pi_*(Q_0S^0) \rightarrow H_*(Q_0S^0).$$

Here and throughout this note, homology is taken with coefficients in \mathbb{F}_2 , the field of two elements. The long-standing conjecture on spherical classes reads as follows.

Conjecture 1.1. There are no spherical classes in Q_0S^0 , except the elements of Hopf invariant 1 and those of Kervaire invariant 1.

An algebraic version of this problem goes as follows. Let V_k be a k -dimensional vector space over \mathbb{F}_2 . Then, the polynomial algebra in k variables $P_k = H^*(BV_k)$ is a module over both the Steenrod algebra \mathcal{A} and the general linear group $GL_k = GL(k, \mathbb{F}_2)$. J. E. Lannes and S. Zarati constructed homomorphisms

$$\varphi_k : Ext_{\mathcal{A}}^{k, k+i}(\mathbb{F}_2, \mathbb{F}_2) \rightarrow (\mathbb{F}_2 \otimes_{\mathcal{A}} P_k^{GL_k})_i^*$$

(see [6]) and have shown that these maps correspond to an associated graded of the Hurewicz homomorphism. The proof of this assertion is unpublished, but it is sketched by J. E. Lannes [5] and by P. G. Goerss [1]. The Hopf invariant 1 and the Kervaire invariant 1 classes are respectively represented by certain permanent cycles in $Ext_{\mathcal{A}}^{1,*}(\mathbb{F}_2, \mathbb{F}_2)$ and $Ext_{\mathcal{A}}^{2,*}(\mathbb{F}_2, \mathbb{F}_2)$, on which φ_1 and φ_2 are non-zero. Therefore, we are led to the following conjecture.

Conjecture 1.2. $\varphi_k = 0$ in any positive stem i for $k > 2$.

In the introduction of the article [2] we are mistaken in asserting that Lannes and Zarati’s work shows that Conjecture 1.2 implies Conjecture 1.1. This comes from the usual problem with spectral sequences: if an element maps to an element of higher filtration, then in the associated graded it will map to 0. Thus φ_k could be 0 even if H is not. Of course, if Conjecture 1.2 were false on a permanent cycle, then Conjecture 1.1 would also be false.

Apart from this, all of our results and proofs in the article [2] are correct.

This correction also applies to our papers [4] (joint with F. P. Peterson) and [3].

The author is grateful to Nick Kuhn for pointing out the above misunderstanding.

Received by the editors February 4, 2003.

2000 *Mathematics Subject Classification*. Primary 55P47, 55Q45, 55S10, 55T15.

REFERENCES

1. P. G. Goerss, *Unstable projectives and stable Ext: with applications*, Proc. London Math. Soc. **53** (1986), 539–561. MR **88d**:55011
2. N. H. V. Hung, *Spherical classes and the algebraic transfer*, Trans. Amer. Math. Soc. **349** (1997), 3893–3910. MR **98e**:55020
3. N. H. V. Hung, *The weak conjecture on spherical classes*, Math. Zeit. **231** (1999), 727–743. MR **2000g**:55019
4. N. H. V. Hung and F. P. Peterson, *Spherical classes and the Dickson algebra*, Math. Proc. Camb. Phil. Soc. **124** (1998), 253–264. MR **99i**:55021
5. J. Lannes, *Sur le n -dual du n -ème spectre de Brown-Gitler*, Math. Zeit. **199** (1988), 29–42. MR **89h**:55020
6. J. Lannes and S. Zarati, *Sur les foncteurs dérivés de la déstabilisation*, Math. Zeit. **194** (1987), 25–59. MR **88j**:55014

DEPARTMENT OF MATHEMATICS, VIETNAM NATIONAL UNIVERSITY, HANOI, 334 NGUYỄN TRÃI STREET, HANOI, VIETNAM

E-mail address: nhvhung@vnu.edu.vn

Editorial Information

To be published in the *Transactions*, a paper must be correct, new, nontrivial, and significant. Further, it must be well written and of interest to a substantial number of mathematicians. Piecemeal results, such as an inconclusive step toward an unproved major theorem or a minor variation on a known result, are in general not acceptable for publication.

Papers submitted to the *Transactions* should exceed 10 published journal pages in length. Shorter papers may be submitted to the *Proceedings of the American Mathematical Society*. Published pages are the same size as those generated in the style files provided for \AA MS-LAT\AA EX or \AA MS-TE\AA X .

As of May 31, 2003, the backlog for this journal was approximately 5 issues. This estimate is the result of dividing the number of manuscripts for this journal in the Providence office that have not yet gone to the printer on the above date by the average number of articles per issue over the previous twelve months, reduced by the number of issues published in four months (the time necessary for editing and composing a typical issue). In an effort to make articles available as quickly as possible, articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue.

A Consent to Publish and Copyright Agreement is required before a paper will be published in this journal. After a paper is accepted for publication, the Providence office will send a Consent to Publish and Copyright Agreement to all authors of the paper. By submitting a paper to this journal, authors certify that the results have not been submitted to nor are they under consideration for publication by another journal, conference proceedings, or similar publication.

Information for Authors

Initial submission. Two copies of the paper should be sent directly to the appropriate Editor and the author should keep a copy. *IF an editor is agreeable*, an electronic manuscript prepared in \AA EX or \AA T\AA EX may be submitted by pointing to an appropriate URL on a preprint or e-print server.

The first page must consist of a *descriptive title*, followed by an *abstract* that summarizes the article in language suitable for workers in the general field (algebra, analysis, etc.). The *descriptive title* should be short, but informative; useless or vague phrases such as “some remarks about” or “concerning” should be avoided. The *abstract* should be at least one complete sentence, and at most 300 words. Included with the footnotes to the paper should be the 2000 *Mathematics Subject Classification* representing the primary and secondary subjects of the article. The classifications are accessible from www.ams.org/msc/. The list of classifications is also available in print starting with the 1999 annual index of *Mathematical Reviews*. The Mathematics Subject Classification footnote may be followed by a list of *key words and phrases* describing the subject matter of the article and taken from it. Journal abbreviations used in bibliographies are listed in the latest *Mathematical Reviews* annual index. The series abbreviations are also accessible from www.ams.org/publications/. To help in preparing and verifying references, the AMS offers MR Lookup, a Reference Tool for Linking, at www.ams.org/mrlookup/. When the manuscript is submitted, authors should supply the editor with electronic addresses if available. These will be printed after the postal address at the end of each article.

Electronically prepared manuscripts. The AMS encourages electronically prepared manuscripts, with a strong preference for \AA MS-LAT\AA EX . To this end, the Society has prepared \AA MS-LAT\AA EX author packages for each AMS publication. Author packages include instructions for preparing electronic manuscripts, the *AMS Author Handbook*, samples, and a style file that generates the particular design specifications of that publication series. Articles properly prepared using the \AA MS-LAT\AA EX style file and the $\backslash\text{label}$ and $\backslash\text{ref}$ commands automatically enable extensive intra-document linking to the bibliography and other elements of the article for searching electronically on the Web. Because linking must often be added manually to electronically prepared manuscripts in other forms of \AA EX , using \AA MS-LAT\AA EX also reduces the amount of technical intervention once the files

are received by the AMS. This results in fewer errors in processing and saves the author proofreading time. \AA MS-L\TeX papers also move more efficiently through the production stream, helping to minimize publishing costs.

\AA MS-L\TeX is the highly preferred format of \TeX , but author packages are also available in \AA MS-T\TeX . Those authors who make use of these style files from the beginning of the writing process will further reduce their own efforts. Manuscripts prepared electronically in \LaTeX or plain \TeX are normally not acceptable due to the high amount of technical time required to insure that the file will run properly through the AMS in-house production system. \LaTeX users will find that \AA MS-L\TeX is the same as \LaTeX with additional commands to simplify the typesetting of mathematics, and users of plain \TeX should have the foundation for learning \AA MS-L\TeX .

Authors may retrieve an author package from the AMS website starting from www.ams.org/tex/ or via FTP to [ftp.ams.org](ftp://ftp.ams.org) (login as `anonymous`, enter username as password, and type `cd pub/author-info`). The *AMS Author Handbook* and the *Instruction Manual* are available in PDF format following the author packages link from www.ams.org/tex/. The author package can also be obtained free of charge by sending email to pub@ams.org (Internet) or from the Publication Division, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When requesting an author package, please specify \AA MS-L\TeX or \AA MS-T\TeX , Macintosh or IBM (3.5) format, and the publication in which your paper will appear. Please be sure to include your complete mailing address.

At the time of submission, authors should indicate if the paper has been prepared using \AA MS-L\TeX or \AA MS-T\TeX and provide the Editor with a paper manuscript that matches the electronic manuscript. The final version of the electronic manuscript should be sent to the Providence office immediately after the paper has been accepted for publication. The author should also send the final version of the paper manuscript to the Editor, who will forward a copy to the Providence office. Editors will require authors to send their electronically prepared manuscripts to the Providence office in a timely fashion. Electronically prepared manuscripts can be sent via email to pub-submit@ams.org (Internet) or on diskette to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. When sending a manuscript electronically, please be sure to include a message indicating in which publication the paper has been accepted. No corrections will be accepted electronically. Authors must mark their changes on their proof copies and return them to the Providence office. Complete instructions on how to send files are included in the author package.

Electronic graphics. Comprehensive instructions on preparing graphics are available starting from www.ams.org/jourhtml/authors.html. A few of the major requirements are given here.

Submit files for graphics as EPS (Encapsulated PostScript) files. This includes graphics originated via a graphics application as well as scanned photographs or other computer-generated images. If this is not possible, TIFF files are acceptable as long as they can be opened in Adobe Photoshop or Illustrator. No matter what method was used to produce the graphic, it is necessary to provide a paper copy to the AMS.

Authors using graphics packages for the creation of electronic art should also avoid the use of any lines thinner than 0.5 points in width. Many graphics packages allow the user to specify a "hairline" for a very thin line. Hairlines often look acceptable when proofed on a typical laser printer. However, when produced on a high-resolution laser imagesetter, hairlines become nearly invisible and will be lost entirely in the final printing process.

Screens should be set to values between 15% and 85%. Screens which fall outside of this range are too light or too dark to print correctly. Variations of screens within a graphic should be no less than 10%.

AMS policy on making changes to articles after posting. Articles are posted to the AMS website individually after proof is returned from authors and before appearing in an issue. To preserve the integrity of electronically published articles, once an article is individually posted to the AMS website but not yet in an issue, changes cannot be made in place in the paper. However, an "Added after posting" section may be added to the

paper right before the References when there is a critical error in the content of the paper. The "Added after posting" section gives the author an opportunity to correct this type of critical error before the article is put into an issue for printing and before it is then reposted with the issue. The "Added after posting" section remains a permanent part of the paper. The AMS does not keep author-related information, such as affiliation, current address, and email address, up to date after a paper is initially posted.

Once the article is assigned to an issue, even if the issue has not yet been posted to the AMS website, corrections may be made to the paper by submitting a traditional errata article to the Editor. The errata article will appear in a future print issue and will link back and forth on the web to the original article online.

Secure manuscript tracking on the Web and via email. Authors can track their manuscripts through the AMS journal production process using the personal AMS ID and Article ID printed in the upper right-hand corner of the Consent to Publish form sent to each author who publishes in AMS journals. Access to the tracking system is available from www.ams.org/mstrack/ or via email sent to mstrack-query@ams.org. To access by email, on the subject line of the message simply enter the AMS ID and Article ID. To track more than one manuscript by email, choose one of the Article IDs and enter the AMS ID and the Article ID followed by the word *all* on the subject line. An explanation of each production step is provided on the web through links from the manuscript tracking screen. Questions can be sent to tran-query@ams.org.

T_EX files available. Beginning with the January 1992 issue of the *Bulletin* and the January 1996 issues of *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS*, T_EX files can be downloaded from the AMS website, starting from www.ams.org/journals/. Authors without Web access may request their files at the address given below after the article has been published. For *Bulletin* papers published in 1987 through 1991 and for *Transactions*, *Proceedings*, *Mathematics of Computation*, and the *Journal of the AMS* papers published in 1987 through 1995, T_EX files are available upon request for authors without Web access by sending email to file-request@ams.org or by contacting the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA. The request should include the title of the paper, the name(s) of the author(s), the name of the publication in which the paper has or will appear, and the volume and issue numbers if known. The T_EX file will be sent to the author making the request after the article goes to the printer. If the requestor can receive Internet email, please include the email address to which the file should be sent. Otherwise please indicate a diskette format and postal address to which a disk should be mailed. **Note:** Because T_EX production at the AMS sometimes requires extra fonts and macros that are not yet publicly available, T_EX files cannot be guaranteed to run through the author's version of T_EX without errors. The AMS regrets that it cannot provide support to eliminate such errors in the author's T_EX environment.

Inquiries. Any inquiries concerning a paper that has been accepted for publication that cannot be answered via the manuscript tracking system mentioned above should be sent to tran-query@ams.org or directly to the Electronic Prepress Department, American Mathematical Society, 201 Charles Street, Providence, RI 02904-2294 USA.

Editors

The traditional method of submitting a paper is to send two hard copies to the appropriate editor. Subjects, and the editors associated with them, are listed below.

In principle the Transactions welcomes electronic submissions, and some of the editors, those whose names appear below with an asterisk (*), have indicated that they prefer them. Editors reserve the right to request hard copies after papers have been submitted electronically. Authors are advised to make preliminary inquiries to editors as to whether they are likely to be able to handle submissions in a particular electronic form.

Algebraic geometry, DAN ABRAMOVICH, Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215 USA; e-mail: abramovic@bu.edu

Algebraic topology and cohomology of groups, STEWART PRIDDY, Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730 USA; e-mail: priddy@math.nwu.edu

* **Combinatorics**, SERGEY FOMIN, Department of Mathematics, East Hall, University of Michigan, Ann Arbor, MI 48109-1109 USA; e-mail: fomin@umich.edu

Complex analysis and geometry, D. H. PHONG, Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027-0029 USA; e-mail: phong@math.columbia.edu

* **Differential geometry and global analysis**, LISA C. JEFFREY, Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3; e-mail: jeffrey@math.toronto.edu

Dynamical systems and ergodic theory, ROBERT F. WILLIAMS, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: bob@math.utexas.edu

* **Geometric analysis**, TOBIAS COLDING, Courant Institute, New York University, 251 Mercer Street, New York, NY 10012 USA; e-mail: colding@cims.nyu.edu

Harmonic analysis, ALEXANDER NAGEL, Department of Mathematics, University of Wisconsin, 480 Lincoln Drive, Madison, WI 53706-1313 USA; e-mail: nagel@math.wisc.edu

Harmonic analysis, representation theory, and Lie theory, ROBERT J. STANTON, Department of Mathematics, Ohio State University, 231 West 18th Avenue, Columbus, OH 43210-1174 USA; e-mail: stanton@math.ohio-state.edu

Number theory, HAROLD G. DIAMOND, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801-2917 USA; e-mail: diamond@math.uiuc.edu

* **Ordinary differential equations, partial differential equations, and applied mathematics**, PETER W. BATES, Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027 USA; e-mail: bates@math.msu.edu

* **Partial differential equations**, PATRICIA E. BAUMAN, Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395 USA; e-mail: bauman@math.purdue.edu

* **Probability and statistics**, KRZYSZTOF BURDZY, Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350 USA; e-mail: burdzy@math.washington.edu

* **Real analysis and partial differential equations**, DANIEL TATARU, Department of Mathematics, University of California, Berkeley, CA 94720 USA; e-mail: tataru@math.berkeley.edu

All other communications to the editors should be addressed to the Managing Editor, WILLIAM BECKNER, Department of Mathematics, University of Texas, Austin, TX 78712-1082 USA; e-mail: beckner@math.utexas.edu

MEMOIRS OF THE AMERICAN MATHEMATICAL SOCIETY

Memoirs is devoted to research in pure and applied mathematics of the same nature as *Transactions*. An issue consists of one or more separately bound research tracts for which the authors provide reproduction copy. Papers intended for *Memoirs* should normally be at least 80 pages in length. *Memoirs* has the same editorial committee as *Transactions*; so such papers should be addressed to one of the editors listed above.

(Continued from back cover)

Toru Ohmoto, Osamu Saeki, and Kazuhiro Sakuma , Self-intersection class for singularities and its application to fold maps	3825
A. Ülger , Erratum to “Arens regularity of the algebra $A \hat{\otimes} B$ ”	3839
Nguyễn H. V. Hưng , Erratum to “Spherical classes and the algebraic transfer”	3841

Heinz H. Bauschke, Frank Deutsch, Hein Hundal, and Sung-Ho Park , Accelerating the convergence of the method of alternating projections	3433
Sunghan Bae, Ernst-Ulrich Gekeler, Pyung-Lyun Kang, and Linsheng Yin , Anderson's double complex and gamma monomials for rational function fields	3463
Lucia Caporaso , Remarks about uniform boundedness of rational points over function fields	3475
Thomas Keilen , Irreducibility of equisingular families of curves	3485
L. Brandolini, A. Iosevich, and G. Travaglini , Planar convex bodies, Fourier transform, lattice points, and irregularities of distribution	3513
Shangbin Cui and Avner Friedman , A free boundary problem for a singular system of differential equations: An application to a model of tumor growth	3537
José García-Cuerva, José Manuel Marco, and Javier Parcet , Sharp Fourier type and cotype with respect to compact semisimple Lie groups	3591
J. Rosický and W. Tholen , Left-determined model categories and universal homotopy theories	3611
Christian Henriksen , The combinatorial rigidity conjecture is false for cubic polynomials	3625
Leo T. Butler , Zero entropy, non-integrable geodesic flows and a non-commutative rotation vector	3641
Carlos Sancho de Salas , Complete homogeneous varieties: Structure and classification	3651
Neil O'Connell , A path-transformation for random walks and the Robinson-Schensted correspondence	3669
Takae Tsuji , On the Iwasawa λ -invariants of real abelian fields	3699
Xiaoxiang Jiao and Jiagui Peng , Pseudo-holomorphic curves in complex Grassmann manifolds	3715
Vassilis G. Papanicolaou , The periodic Euler-Bernoulli equation	3727
Francisco Jesús Castro-Jiménez and Nobuki Takayama , Singularities of the hypergeometric system associated with a monomial curve	3761
Jason P. Bell and Stanley N. Burris , Asymptotics for logical limit laws: When the growth of the components is in an RT class	3777
Michael E. Hoffman , Combinatorics of rooted trees and Hopf algebras	3795
Mahuya Datta , Connections with prescribed first Pontrjagin form	3813

(Continued on inside back cover)





American Mathematical Society. Transactions
v.355, no.7-9 ; 2003

MATHEMATICAL SCIENCES LIBRARY
UNIVERSITY OF



GretagMacbeth™ ColorChecker Color Rendition Chart